CONSISTENT EXPLICIT STAGGERED SCHEMES FOR COMPRESSIBLE FLOWS

PART II: THE EULER EQUATIONS.

R. HERBIN *, J.-C. LATCHÉ †, AND TT. NGUYEN ‡

May 7, 2013

Abstract. In this paper, we build and analyze the stability and consistency of an explicit scheme for the Euler equations. This scheme is based on staggered space discretizations, with an upwinding performed with respect to the material velocity only. The pressure gradient is defined as the transpose of the natural velocity divergence, and is thus centered. The energy equation which is solved is the internal energy balance, which offers two main advantages: first, we avoid the space discretization of the total energy, the expression of which involves cell-centered and face-centered variables; second, the discretization ensures by construction the positivity of the internal energy, under a CFL condition. However, since this scheme does not use the original (total) energy conservative equation, in order to obtain correct weak solutions (in particular, with shocks satisfying the Rankine-Hugoniot conditions), we need to introduce corrective terms in the internal energy balance. These corrective terms are found by deriving a discrete kinetic energy balance, observing that this relation contains residual terms which do not tend to zero (at least, under reasonable stability assumptions) and, finally, compensating them in the discrete internal energy balance. It is then shown in the 1D case, that, if the scheme converges, the limit is indeed a weak solution. Finally, we present numerical results which confort this theory.

Key words. finite volumes, finite elements, staggered discretizations, Euler equations, compressible flows, analysis.

AMS subject classifications. 65M12

1. Introduction. We address in this paper the so-called Euler equations, which read:

$$\partial_t \rho + \mathrm{div}(\rho\, \boldsymbol{u}) = 0, \tag{1.1a}$$

$$\partial_t (\rho\, \boldsymbol{u}) + \mathrm{div}(\rho\, \boldsymbol{u} \otimes \boldsymbol{u}) + \boldsymbol{\nabla} p = 0, \tag{1.1b}$$

$$\partial_t (\rho\, E) + \mathrm{div}(\rho\, E\, \boldsymbol{u}) + \mathrm{div}(p\, \boldsymbol{u}) = 0, \tag{1.1c}$$

$$p = (\gamma - 1)\, \rho\, e, \qquad E = \frac{1}{2}|\boldsymbol{u}|^2 + e, \tag{1.1d}$$

where t stands for the time, ρ, $\boldsymbol{u}$, p, E and e are the density, velocity, pressure, total energy and internal energy respectively, and $\gamma > 1$ is a coefficient specific to the considered fluid. The problem is supposed to be posed over $\Omega \times (0, T)$, where Ω is an open bounded connected subset of $\mathbb{R}^d$, $1 \leq d \leq 3$, and $(0, T)$ is a finite time interval.

System (1.1) is complemented by initial conditions for ρ, e and $\boldsymbol{u}$, denoted by ρ_0, e_0 and $\boldsymbol{u}_0$ respectively, with $\rho_0 > 0$ and $e_0 > 0$, and by a boundary condition which we suppose to be $\boldsymbol{u} \cdot \boldsymbol{n} = 0$ at any time and *a.e.* on $\partial\Omega$, where $\boldsymbol{n}$ stands for the normal vector to the boundary.

Let us suppose that the solution is regular, and let E_k be the kinetic energy, defined by $E_k = \frac{1}{2}\,|\boldsymbol{u}|^2$. Taking the inner product of (1.1b) by $\boldsymbol{u}$ yields, after formal compositions of partial derivatives and using the mass balance (1.1a):

$$\partial_t(\rho E_k) + \mathrm{div}\big(\rho\, E_k\, \boldsymbol{u}\big) + \boldsymbol{\nabla} p \cdot \boldsymbol{u} = 0. \tag{1.2}$$

This relation is referred to as the kinetic energy balance. Substracting this relation from the total energy balance (1.1c), we obtain the internal energy balance equation:

$$\partial_t(\rho e) + \mathrm{div}(\rho e \boldsymbol{u}) + p\, \mathrm{div}\boldsymbol{u} = 0. \tag{1.3}$$

Since,

*Université d'Aix-Marseille (herbin@cmi.univ-mrs.fr)
†Institut de Radioprotection et de Sûreté Nucléaire (IRSN) (jean-claude.latche@irsn.fr)
‡Institut de Radioprotection et de Sûreté Nucléaire (IRSN) (tan-trung.nguyen@irsn.fr)

- thanks to the mass balance equation, the first two terms in the left-hand side of (1.3) may be recast as a transport operator: $\partial_t(\rho e)+\mathrm{div}(\rho e\boldsymbol{u}) = \rho\,[\partial_t e+\boldsymbol{u}\cdot\boldsymbol{\nabla} e]$,
- and, from the equation of state, the pressure vanishes when $e = 0$,

this equation implies, if $e \geq 0$ at $t = 0$ and with suitable boundary conditions, that e remains non-negative at all times.

The objective pursued in this work is to develop and study, from a theoretical point of view, an explicit scheme for the solution of (1.1). More precisely, we intend to build an explicit variant of pressure correction schemes that were developed and studied recently in the framework of the simulation of compressible flows at all speeds [13, 11, 12], and implemented in the industrial open-source computer code ISIS [16]. Indeed, our initial motivation was to provide in the same software an efficient alternative of these schemes for quickly varying unsteady flows, with a characteristic Mach number in the range or greater than 1. In order to remain stable in the incompressible limit, the starting-point algorithms are based on (*inf-sup* stable) staggered finite volume or finite element discretizations, and the present scheme thus also relies on these space approximations. In our approach, the upwinding techniques which are implemented for stability reasons are performed for each equation separately and with respect to the material velocity only. This is in contradiction with the most common strategy adopted for hyperbolic systems, where upwinding is built from the wave structure of the system (see *e.g.* [21, 6, 2] for surveys). However, it yields algorithms which are used in practice (see *e.g.* the so-called AUSM family of schemes [19, 18]), because of their generality (a closed-form solution of Riemann problems is not needed), theirease of implementation and their efficiency, thanks to an easy construction of the fluxes at the cell faces.

Another salient feature of the proposed scheme is that we discretize the internal energy balance (1.3) instead of the total energy balance (1.1c); this presents two advantages:

- first, it avoids the space discretization of the total energy, which is rather unnatural for staggered schemes since the degrees of freedom for the velocity and the scalar variables are not collocated,
- second, by a suitable discretization of (1.3), we obtain a scheme which ensures, "by construction", the positivity of the internal energy.

However, for solutions with shocks, Equation (1.3) is not equivalent to (1.1c); more precisely speaking, at the locations of shocks, positive measures should appear, at the right-hand side of Equation (1.3). Discretizing (1.3) instead of (1.1c) may thus yield a scheme which does not compute the correct weak discontinuous solutions; in particular, the numerical solutions may present (smeared) shocks which do not satisfy the Rankine-Hugoniot conditions associated to (1.1c). The essential result of this paper is to provide solutions to circumvent this problem. To this purpose, we closely mimic the above performed formal computation:

- we establish a discrete equivalent of the kinetic energy balance (1.3), and remark that the residual terms at the right-hand side do no tend to zero with the space and time steps (they are the discrete manifestations of the above mentioned measures),
- we thus compensate these residual terms by corrective terms in the internal energy balance.

We provide a theoretical justification of this process by showing that, in the 1D case, if the scheme is stable and converges to a limit (in a sense to be defined), this limit satisfies a weak form of (1.1c) which implies the correct Rankine-Hugoniot conditions. This paper is structured as follows. We begin with the presentation of the space discretization (Section 2), then the scheme is given (Section 3), and we derive the discrete kinetic and elastic potential balances satisfied by its solutions (Section 4). Next section is dedicated to the proof, in 1D, of the consistency of the scheme (Section 5). We then present some numerical tests, to assess the behaviour of the algorithm (Section 6).

The present work is an extension of a companion paper devoted to the barotropic Euler equations [15]. However, we provide here a self-consistent presentation of the discretization and the scheme, and only refer to [15] for the proof of some theoretical results.

2. Meshes and unknowns. In this section, we focus on the discretization of a multi-dimensional domain (*i.e.* $d = 2$ or $d = 3$); the extension to the one-dimensional case is straightforward (see Section 5).

Let $\mathcal{M}$ be a mesh of the domain Ω, supposed to be regular in the usual sense of the finite element literature (*e.g.* [3]). The cells of the mesh are assumed to be:

- for a general domain Ω, either non-degenerate quadrilaterals ($d = 2$) or hexahedra ($d = 3$) or simplices, both type of cells being possibly combined in a same mesh,
- for a domain the boundaries of which are hyperplanes normal to a coordinate axis, rectangles ($d = 2$) or rectangular parallelepipeds ($d = 3$) (the faces of which, of course, are then also necessarily normal to a coordinate axis).

By $\mathcal{E}$ and $\mathcal{E}(K)$ we denote the set of all $(d-1)$-faces σ of the mesh and of the element $K \in \mathcal{M}$ respectively. The set of faces included in the boundary of Ω is denoted by $\mathcal{E}_{\rm ext}$ and the set of internal faces (*i.e.* $\mathcal{E} \setminus \mathcal{E}_{\rm ext}$) is denoted by $\mathcal{E}_{\rm int}$; a face $\sigma \in \mathcal{E}_{\rm int}$ separating the cells K and L is denoted by $\sigma = K|L$. The outward normal vector to a face σ of K is denoted by $\boldsymbol{n}_{K,\sigma}$. For $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}$, we denote by $|K|$ the measure of K and by $|\sigma|$ the $(d-1)$-measure of the face σ. For $1 \leq i \leq d$, we denote by $\mathcal{E}^{(i)} \subset \mathcal{E}$ and $\mathcal{E}^{(i)}_{\rm ext} \subset \mathcal{E}_{\rm ext}$ the subset of the faces of $\mathcal{E}$ and $\mathcal{E}_{\rm ext}$ respectively which are perpendicular to the i^{th} unit vector of the canonical basis of $\mathbb{R}^d$.

The space discretization is staggered, using either the Marker-And Cell (MAC) scheme [10, 9], or nonconforming low-order finite element approximations, namely the Rannacher and Turek element (RT) [20] for quadrilateral or hexahedric meshes, or the lowest degree Crouzeix-Raviart element (CR) [4] for simplicial meshes.

For all these space discretizations, the degrees of freedom for the pressure, the density and the internal energy (*i.e.* the discrete pressure, density and internal energy unknowns) are associated to the cells of the mesh $\mathcal{M}$, and are denoted by:

$$\big\{p_K,\ \rho_K,\ e_K,\ K \in \mathcal{M}\big\}.$$

Let us then turn to the degrees of freedom for the velocity (*i.e.* the discrete velocity unknowns).

- **Rannacher-Turek** or **Crouzeix-Raviart** discretizations – The degrees of freedom for the velocity components are located at the center of the faces of the mesh, and we choose the version of the element where they represent the average of the velocity through a face. The set of degrees of freedom reads:

$$\{u_{\sigma,i},\ \sigma \in \mathcal{E},\ 1 \leq i \leq d\}.$$

- **MAC** discretization – The degrees of freedom for the i^{th} component of the velocity are defined at the centre of the faces $\sigma \in \mathcal{E}^{(i)}$, so the whole set of discrete velocity unknowns reads:

$$\big\{u_{\sigma,i},\ \sigma \in \mathcal{E}^{(i)},\ 1 \leq i \leq d\big\}.$$

We now introduce a dual mesh, which will be used for the finite volume approximation of the time derivative and convection terms in the momentum balance equation.

- **Rannacher-Turek** or **Crouzeix-Raviart** discretizations – For the RT or CR discretizations, the dual mesh is the same for all the velocity components. When $K \in \mathcal{M}$ is a simplex, a rectangle or a cuboid, for $\sigma \in \mathcal{E}(K)$, we define $D_{K,\sigma}$ as the cone with basis σ and with vertex the mass center of K (see Figure 2.1). We thus obtain a partition of K in m sub-volumes, where m is the number of

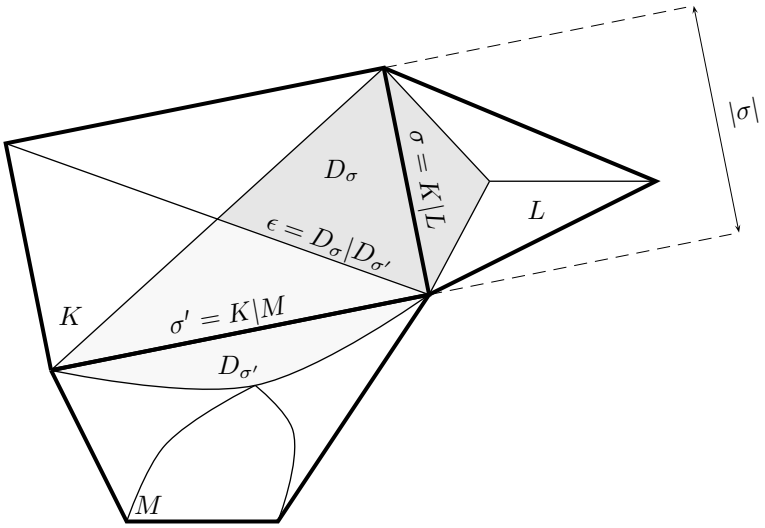

FIG. 2.1. *Primal and dual meshes for the Rannacher-Turek and Crouzeix-Raviart elements.*

faces of the mesh, each sub-volume having the same measure $|D_{K,\sigma}| = |K|/m$. We extend this definition to general quadrangles and hexahedra, by supposing that we have built a partition still of equal-volume sub-cells, and with the same connectivities. The volume $D_{K,\sigma}$ is referred to as the half-diamond cell associated to K and σ.
For $\sigma \in \mathcal{E}_{\rm int}$, $\sigma = K|L$, we now define the diamond cell D_σ associated to σ by $D_\sigma = D_{K,\sigma} \cup D_{L,\sigma}$; for an external face $\sigma \in \mathcal{E}_{\rm ext} \cap \mathcal{E}(K)$, D_σ is just the same volume as $D_{K,\sigma}$.

- **MAC** discretization – For the MAC scheme, the dual mesh depends on the component of the velocity. For each component, the MAC dual mesh only differs from the RT or CR dual mesh by the choice of the half-diamond cell, which, for $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}(K)$, is now the rectangle or rectangular parallelepiped of basis σ and of measure $|D_{K,\sigma}| = |K|/2$.

We denote by $|D_\sigma|$ the measure of the dual cell D_σ, and by $\epsilon = D_\sigma|D_{\sigma'}$ the face separating two diamond cells D_σ and $D_{\sigma'}$. The set of the faces of a dual cell D_σ is denoted by $\tilde{\mathcal{E}}(D_\sigma)$.

Finally, we need to deal with the impermeability (*i.e.* $\boldsymbol{u} \cdot \boldsymbol{n} = 0$) boundary condition. Since the velocity unknowns lie on the boundary (and not inside the cells), these conditions are taken into account in the definition of the discrete spaces. To avoid technicalities in the expression of the schemes, we suppose throughout this paper that the boundary is *a.e.* normal to a coordinate axis, (even in the case of the RT or CR discretizations), which allows to simply set to zero the corresponding velocity unknowns:

$$\text{for } i = 1, \ldots, d, \ \forall \sigma \in \mathcal{E}_{\rm ext}^{(i)}, \qquad u_{\sigma,i} = 0. \tag{2.1}$$

Therefore, there are no degrees of freedom for the velocity on the boundary for the MAC scheme, and there are only $d-1$ degrees of freedom on each boundary face for the CR and RT discretizations, which depend on the orientation of the face. In order to be able to write a unique expression of the discrete equations for both MAC and CR/RT schemes, we introduce the set of faces $\mathcal{E}_{\mathcal{S}}^{(i)}$ associated to the degrees of freedom of each component of the velocity ($\mathcal{S}$ stands for "scheme"):

$$\mathcal{E}_{\mathcal{S}}^{(i)} = \left| \begin{array}{l} \mathcal{E}^{(i)} \setminus \mathcal{E}_{\rm ext}^{(i)} \text{ for the MAC scheme,} \\ \mathcal{E} \setminus \mathcal{E}_{\rm ext}^{(i)} \text{ for the CR or RT schemes.} \end{array} \right.$$

For both schemes, we define $\tilde{\mathcal{E}}^{(i)}$, for $1 \leq i \leq d$, as the set of faces of the dual mesh associated to the i^{th} component of the velocity. For the RT or CR discretizations, the sets $\tilde{\mathcal{E}}^{(i)}$ does not depend on the component (*i.e.* of i), up to the elimination of some unknowns (and so some dual cells and, finally, some external faces) to take the boundary conditions into account. For the MAC scheme, $\tilde{\mathcal{E}}^{(i)}$ depends on i; note that each face of $\tilde{\mathcal{E}}^{(i)}$ is perpendicular to a unit vector of the canonical basis of $\mathbb{R}^d$, but not necessarily to the i^{th} one.

General domains can be addressed (of course, with the CR or RT discretizations) by redefining, through linear combinations, the degrees of freedom at the external faces, so as to introduce the normal velocity as a new degree of freedom.

3. The scheme. Let us consider a partition $0 = t_0 < t_1 < \ldots < t_N = T$ of the time interval $(0, T)$, which we suppose uniform for the sake of simplicity, and let $\delta t = t_{n+1} - t_n$ for $n = 0, 1, \ldots, N-1$ be the (constant) time step. We consider an explicit-in-time scheme, which reads in its fully discrete form, for $0 \leq n \leq N-1$:

$$\forall K \in \mathcal{M},\ \frac{|K|}{\delta t}(\rho_K^{n+1} - \rho_K^n) + \sum_{\sigma \in \mathcal{E}(K)} F_{K,\sigma}^n = 0, \tag{3.1a}$$

$$\forall K \in \mathcal{M},\ \frac{|K|}{\delta t}(\rho_K^{n+1} e_K^{n+1} - \rho_K^n e_K^n) + \sum_{\sigma \in \mathcal{E}(K)} F_{K,\sigma}^n e_\sigma^n + |K|\ p_K^n\ (\mathrm{div}\boldsymbol{u})_K^n = S_K^n, \tag{3.1b}$$

$$\forall K \in \mathcal{M},\ p_K^{n+1} = (\gamma - 1)\ \rho_K^{n+1}\ e_K^{n+1}, \tag{3.1c}$$

For $1 \leq i \leq d,\ \forall \sigma \in \mathcal{E}_{\mathcal{S}}^{(i)}$,

$$\frac{|D_\sigma|}{\delta t}(\rho_{D_\sigma}^{n+1} u_{\sigma,i}^{n+1} - \rho_{D_\sigma}^n u_{\sigma,i}^n) + \sum_{\epsilon \in \tilde{\mathcal{E}}(D_\sigma)} F_{\sigma,\epsilon}^n u_{\epsilon,i}^n + |D_\sigma|\,(\boldsymbol{\nabla} p)_{\sigma,i}^{n+1} = 0, \tag{3.1d}$$

where the terms introduced for each discrete equation are defined hereafter.

Equation (3.1a) is obtained by the discretization of the mass balance equation (1.1a) over the primal mesh, and $F_{K,\sigma}^n$ stands for the mass flux across σ outward K, which, because of the impermeability condition, vanishes on external faces and is given on the internal faces by:

$$\forall \sigma = K|L \in \mathcal{E}_{\mathrm{int}}, \qquad F_{K,\sigma}^n = |\sigma|\ \rho_\sigma^n\ u_{K,\sigma}^n, \tag{3.2}$$

where $u_{K,\sigma}^n$ is an approximation of the normal velocity to the face σ outward K. This latter quantity is defined by:

$$u_{K,\sigma}^n = \begin{cases} u_{\sigma,i}^n\ \boldsymbol{e}^{(i)} \cdot \boldsymbol{n}_{K,\sigma} \text{ for } \sigma \in \mathcal{E}^{(i)} \text{ in the MAC case,} \\ \boldsymbol{u}_\sigma^n \cdot \boldsymbol{n}_{K,\sigma} \text{ in the CR and RT cases,} \end{cases} \tag{3.3}$$

where $\boldsymbol{e}^{(i)}$ denotes the i-th vector of the orthonormal basis of $\mathbb{R}^d$. The density at the face $\sigma = K|L$ is approximated by the upwind technique:

$$\rho_\sigma^n = \begin{cases} \rho_K^n & \text{if } u_{K,\sigma}^n \geq 0, \\ \rho_L^n & \text{otherwise.} \end{cases} \tag{3.4}$$

We now turn to the discrete momentum balance (3.1d), which is obtained by discretizing the momentum balance equation (1.1b) on the dual cells associated to the faces of the mesh. The first task is to define the values $\rho_{D_\sigma}^{n+1}$ and $\rho_{D_\sigma}^n$, which approximate the density over the dual cell D_σ at time t^{n+1} and t^n respectively, and the discrete mass flux through the dual face ϵ outward D_σ, denoted by $F_{\sigma,\epsilon}^n$; the

guideline for their construction is that a finite volume discretization of the mass balance equation over the diamond cells, of the form

$$\forall \sigma \in \mathcal{E}, \qquad \frac{|D_\sigma|}{\delta t}\,(\rho_{D_\sigma}^{n+1} - \rho_{D_\sigma}^{n}) + \sum_{\epsilon \in \tilde{\mathcal{E}}(D_\sigma)} F_{\sigma,\epsilon}^{n} = 0, \tag{3.5}$$

must hold in order to be able to derive a discrete kinetic energy balance (see Section 4 below). The density on the dual cells is given by the following weighted average:

$$\text{for } \sigma = K|L \in \mathcal{E}_{\rm int}, \text{ for } k = n \text{ and } k = n+1,$$
$$|D_\sigma|\ \rho_{D_\sigma}^{k} = |D_{K,\sigma}|\ \rho_K^k + |D_{L,\sigma}|\ \rho_L^k. \tag{3.6}$$

For the MAC scheme, the flux on a dual face which is located on two primal faces is the mean value of the sum of fluxes on the two primal faces, and the flux of a dual face located between two primal faces is again the mean value of the sum of fluxes on the two primal faces [14]. In the case of the CR and RT schemes, for a dual face ϵ included in the primal cell K, this flux is computed as a linear combination (with constant coefficients, *i.e.* independent of the cell) of the mass fluxes through the faces of K, *i.e.* the quantities $(F_{K,\sigma}^n)_{\sigma \in \mathcal{E}(K)}$ appearing in the discrete mass balance (3.1a). We refer to [1, 5] for a detailed construction of this approximation. Let us remark that a dual face lying on the boundary is then also a primal face, and the flux across this face is zero. Therefore, the values $u_{\epsilon,i}^n$ are only needed at the internal dual faces, and we make the upwind choice for their discretization:

$$\text{for } \epsilon = D_\sigma|D_{\sigma'}, \quad u_{\epsilon,i}^n = \begin{cases} u_{\sigma,i}^n & \text{if } F_{\sigma,\epsilon}^n \geq 0, \\ u_{\sigma',i}^n & \text{otherwise.} \end{cases} \tag{3.7}$$

The last term $(\boldsymbol{\nabla} p)_{\sigma,i}^{n+1}$ stands for the i-th component of the discrete pressure gradient at the face σ. The gradient operator is built as the transpose of the discrete operator for the divergence of the velocity, the discretization of which is based on the primal mesh. Let us denote the divergence of $\boldsymbol{u}^{n+1}$ over $K \in \mathcal{M}$ by $(\mathrm{div}\boldsymbol{u})_K^{n+1}$; its natural approximation reads:

$$\text{for } K \in \mathcal{M}, \qquad (\mathrm{div}\boldsymbol{u})_K^{n+1} = \frac{1}{|K|} \sum_{\sigma \in \mathcal{E}(K)} |\sigma|\ u_{K,\sigma}^{n+1}. \tag{3.8}$$

Consequently, the components of the pressure gradient are given by:

$$\text{for } \sigma = K|L \in \mathcal{E}_{\rm int}, \qquad (\boldsymbol{\nabla} p)_{\sigma,i}^{n+1} = \frac{|\sigma|}{|D_\sigma|}\,(p_L^{n+1} - p_K^{n+1})\ \boldsymbol{n}_{K,\sigma} \cdot \boldsymbol{e}^{(i)}, \tag{3.9}$$

this expression being derived thanks to the following duality relation with respect to the L^2 inner product:

$$\sum_{K \in \mathcal{M}} |K|\ p_K^{n+1}\ (\mathrm{div}\boldsymbol{u})_K^{n+1} + \sum_{i=1}^{d} \sum_{\sigma \in \mathcal{E}_S^{(i)}} |D_\sigma|\ u_{\sigma,i}^{n+1}\ (\boldsymbol{\nabla} p)_{\sigma,i}^{n+1} = 0. \tag{3.10}$$

Note that, because of the impermeability boundary conditions, the discrete gradient is not defined at the external faces.

Equation (3.1b) is an approximation of the internal energy balance over the primal cell K. The positivity of the convection operator is ensured if we use an upwinding technique for this term [17]:

$$\text{for } \sigma = K|L \in \mathcal{E}_{\rm int}, \qquad e_\sigma^n = \begin{cases} e_K^n & \text{if } F_{K,\sigma}^n \geq 0, \\ e_L^n & \text{otherwise.} \end{cases}$$

The discrete divergence of the velocity, $(\mathrm{div}\boldsymbol{u})_K^n$, is defined by (3.8). The right-hand side, S_K^n, is derived using consistency arguments in the next section; at the first time step, it is simply set to zero:

$$\forall K \in \mathcal{M}, \qquad S_K^0 = 0.$$

Finally, the initial approximations for ρ, e and $\boldsymbol{u}$ are given by the average of the initial conditions ρ_0 and e_0 on the primal cells and of $\boldsymbol{u}_0$ on the dual cells:

$$\begin{aligned} &\forall K \in \mathcal{M}, \quad \rho_K^0 = \frac{1}{|K|}\int_K \rho_0(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}, \quad \text{and } e_K^0 = \frac{1}{|K|}\int_K e_0(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}, \\ &\text{for } 1 \le i \le d,\ \forall \sigma \in \mathcal{E}_{\mathcal{S}}^{(i)}, \quad u_{\sigma,i}^0 = \frac{1}{|D_\sigma|}\int_{D_\sigma} (\boldsymbol{u}_0(\boldsymbol{x}))_i\,\mathrm{d}\boldsymbol{x}. \end{aligned} \tag{3.11}$$

The following positivity result is a classical consequence of the upwind choice in the mass balance equation.

LEMMA 3.1 (Positivity of the density). *Let ρ^0 be given by* (3.11). *Then, since ρ_0 is assumed to be a positive function, $\rho^0 > 0$ and, under the* CFL *condition:*

$$\delta t \le \frac{|K|}{\sum_{\sigma\in\mathcal{E}(K)} |\sigma|\ (u_{K,\sigma}^n)^+}, \quad \forall K \in \mathcal{M},\ \textit{for } 0 \le n \le N-1, \tag{3.12}$$

where, for $a \in \mathbb{R}$, $a^+ \ge 0$ is defined by $a^+ = \max(a, 0)$, the solution to the scheme satisfies $\rho^n > 0$, for $1 \le n \le N$.

4. Discrete kinetic energy balance and corrective source terms. Equation (4.1) below is a discrete analogue of the kinetic energy balance equation (1.2), with an upwind discretization of the convection term. Its proof may be found in [15] (Lemma 4.1).

LEMMA 4.1 (Discrete kinetic energy balance).
A solution to the system (3.1) *satisfies the following equality, for $1 \le i \le d$, $\sigma \in \mathcal{E}_{\mathcal{S}}^{(i)}$ and $0 \le n \le N-1$:*

$$\begin{aligned} \frac{1}{2}\frac{|D_\sigma|}{\delta t}\left[\rho_{D_\sigma}^{n+1}(u_{\sigma,i}^{n+1})^2 - \rho_{D_\sigma}^n (u_{\sigma,i}^n)^2\right] + \frac{1}{2}\sum_{\epsilon\in\tilde{\mathcal{E}}(D_\sigma)} F_{\sigma,\epsilon}^n\ (u_{\epsilon,i}^n)^2 \\ + |D_\sigma|\,(\boldsymbol{\nabla} p)_{\sigma,i}^{n+1}\ u_{\sigma,i}^{n+1} = -R_{\sigma,i}^{n+1}, \end{aligned} \tag{4.1}$$

with:

$$\begin{aligned} R_{\sigma,i}^{n+1} = \frac{1}{2}\frac{|D_\sigma|}{\delta t}\rho_{D_\sigma}^{n+1}(u_{\sigma,i}^{n+1} - u_{\sigma,i}^n)^2 + \frac{1}{2}\sum_{\epsilon = D_\sigma | D_{\sigma'} \in \tilde{\mathcal{E}}(D_\sigma)} (F_{\sigma,\epsilon}^n)^-(u_{\sigma',i}^n - u_{\sigma,i}^n)^2 \\ - \sum_{\epsilon = D_\sigma | D_{\sigma'} \in \tilde{\mathcal{E}}(D_\sigma)} (F_{\sigma,\epsilon}^n)^-(u_{\sigma,i}^{n+1} - u_{\sigma,i}^n)\,(u_{\sigma',i}^n - u_{\sigma,i}^n), \end{aligned} \tag{4.2}$$

where, for $a \in \mathbb{R}$, $a^- \ge 0$ is defined by $a^- = -\min(a, 0)$.

The next step is now to define corrective terms in the internal energy balance, with the aim to recover a consistent discretization of the total energy balance. The first idea to do this could be just to sum the (discrete) kinetic energy balance with the internal energy balance: it is indeed possible for a collocated discretization. But here, we face the fact that the kinetic energy balance is associated to the dual mesh, while the internal energy balance is discretized on the primal mesh. The way to circumvent this difficulty is to remark that we do not really need a discrete total energy balance; in fact, we only need to recover (a weak form of) this equation when the mesh and time steps tend to zero. To this purpose, we choose the quantities (S_K^{n+1}) in such a

way as to somewhat compensate the terms $(R_{\sigma,i}^{n+1})$ given by (4.2). For $K \in \mathcal{M}$, we obtain $S_K^{n+1} = \sum_{i=1}^d S_{K,i}^{n+1}$ with:

$$S_{K,i}^{n+1} = \frac{1}{2}\,\rho_K^{n+1} \sum_{\sigma\in\mathcal{E}(K)\cap\mathcal{E}_{\mathcal{S}}^{(i)}} \frac{|D_{K,\sigma}|}{\delta t}\,\left(u_{\sigma,i}^{n+1} - u_{\sigma,i}^n\right)^2$$
$$+ \sum_{\substack{\epsilon\in\tilde{\mathcal{E}}_{\mathcal{S}}^{(i)},\ \epsilon\cap\bar{K}\neq\emptyset,\\ \epsilon = D_\sigma | D_{\sigma'},\, F_{\sigma,\epsilon}^n \leq 0}} \alpha_{K,\epsilon}\,\Big[\frac{|F_{\sigma,\epsilon}^n|}{2}\,(u_{\sigma,i}^n - u_{\sigma',i}^n)^2 + F_{\sigma,\epsilon}^n\,\left(u_{\sigma,i}^{n+1} - u_{\sigma,i}^n\right)\,(u_{\sigma',i}^n - u_{\sigma,i}^n)\Big]. \tag{4.3}$$

The coefficient $\alpha_{K,\epsilon}$ is fixed to 1 if the face ϵ is included in K, and this is the only situation to consider for the RT and CR discretizations. For the MAC scheme, some dual faces are included in the primal cells, but some lie on their boundary; for such a boundary edge ϵ, we denote by $\mathcal{N}_\epsilon$ the set of cells M such that $\bar{M}\cap\epsilon \neq \emptyset$ (the cardinal of this set is always 4, except for boundary edges through which, anyway, the mass flux vanishes), and compute $\alpha_{K,\epsilon}$ by:

$$\alpha_{K,\epsilon} = \frac{|K|}{\sum_{M\in\mathcal{N}_\epsilon} |M|}. \tag{4.4}$$

For a uniform grid, this formula yields $\alpha_{K,\epsilon} = 1/4$.

The expression of the $(S_K^{n+1})_{K\in\mathcal{M}}$ is justified by the passage to the limit in the scheme (for a one-dimensional problem) performed in the next section. We note however here that:

$$\sum_{K\in\mathcal{M}} S_K^{n+1} - \sum_{i=1}^d \sum_{\sigma\in\mathcal{E}_{\mathcal{S}}^{(i)}} R_{\sigma,i}^{n+1} = 0. \tag{4.5}$$

Indeed, the first part of $S_{K,i}^{n+1}$, thanks to the expression (3.6) of the density at the face $\rho_{D_\sigma}^{n+1}$, results from a dispatching of the first part of the residual over the two adjacent cells:

$$\frac{1}{2}\,\frac{|D_\sigma|}{\delta t}\,\rho_{D_\sigma}^{n+1}\,\left(u_{\sigma,i}^{n+1} - u_{\sigma,i}^n\right)^2 =$$
$$\underbrace{\frac{1}{2}\,\frac{|D_{K,\sigma}|}{\delta t}\,\rho_K^{n+1}\,\left(u_{\sigma,i}^{n+1} - u_{\sigma,i}^n\right)^2}_{\text{affected to K}} + \underbrace{\frac{1}{2}\,\frac{|D_{L,\sigma}|}{\delta t}\,\rho_L^{n+1}\,\left(u_{\sigma,i}^{n+1} - u_{\sigma,i}^n\right)^2}_{\text{affected to L}}.$$

The same argument holds for the terms associated to the dual faces, which explains, in particular, the definition of the coefficients $\alpha_{K,\epsilon}$. The scheme thus conserves the integral of the total energy over the computational domain. In the scheme itself, we shall use the term S_K^n rather than S_K^{n+1}, because we want an explicit scheme, but this does not hinder the consitency of the scheme, as shown in the proof of Theorem 5.2. The definition (4.3) of $(S_K^{n+1})_{K\in\mathcal{M}}$ allows to prove that, under a CFL condition, the scheme preserves the positivity of e.

LEMMA 4.2. *Let us suppose that, for $0 \leq n \leq N-1$, for all $K \in \mathcal{M}$ and $\sigma \in \mathcal{E}(K)$, we have:*

$$\delta t \leq \min\Big(\frac{|K|}{\gamma \sum_{\sigma\in\mathcal{E}(K)} |\sigma|\,(u_{K,\sigma}^n)^+}, \frac{|D_{K,\sigma}|\,\rho_K^{n+1}}{\sum_{\epsilon\in\tilde{\mathcal{E}}(D_\sigma),\ \epsilon\cap\bar{K}\neq\emptyset} \alpha_{K,\epsilon}\,(F_{\sigma,\epsilon}^n)^-}\Big). \tag{4.6}$$

Then the internal energy $(e^n)_{1\leq n\leq N}$ given by the scheme (3.1) is positive.

Proof. Let n such that $0 \le n \le N$ be given, and let us assume that $e_K^n \ge 0$ and $S_K^n \ge 0$ for all $K \in \mathcal{M}$. Since, by assumption, $\gamma > 1$, the CFL condition (4.6) implies that the CFL condition (3.12) is satisfied, and by Lemma 3.1 we thus have $\rho_K^n > 0$ and $\rho_K^{n+1} > 0$, for all $K \in \mathcal{M}$. In the internal energy equation (3.1b), let us express the pressure thanks to the equation of state (3.1c) to obtain:

$$\frac{|K|}{\delta t}\rho_K^{n+1}e_K^{n+1} = \Big[\frac{|K|}{\delta t}\rho_K^n - \sum_{\sigma\in\mathcal{E}(K)} (F_{K,\sigma}^n)^+ - (\gamma-1)\,\rho_K^n \sum_{\sigma\in\mathcal{E}(K)} |\sigma|\,(u_{K,\sigma}^n)^+\Big]\, e_K^n + \sum_{\sigma\in\mathcal{E}(K)} (F_{K,\sigma}^n)^- e_L^n + (\gamma-1)\,\rho_K^n\, e_K^n \sum_{\sigma\in\mathcal{E}(K)} |\sigma|\,(u_{K,\sigma}^n)^- + S_K^n. \quad (4.7)$$

Using the fact that, when $u_{K,\sigma}^n \ge 0$, the upwind density at the face is ρ_K^n, we have:

$$(F_{K,\sigma}^n)^+ + (\gamma-1)\,|\sigma|\,\rho_K^n\,(u_{K,\sigma}^n)^+ = \gamma\,|\sigma|\,\rho_K^n\,(u_{K,\sigma}^n)^+,$$

and hence Relation (4.7) reads:

$$\frac{|K|}{\delta t}\rho_K^{n+1}e_K^{n+1} = \Big[\frac{|K|}{\delta t} - \gamma \sum_{\sigma\in\mathcal{E}(K)} |\sigma|\,(u_{K,\sigma}^n)^+\Big]\,\rho_K^n\, e_K^n + \sum_{\sigma\in\mathcal{E}(K)} (F_{K,\sigma}^n)^- e_L^n + (\gamma-1)\,\rho_K^n\, e_K^n \sum_{\sigma\in\mathcal{E}(K)} |\sigma|\,(u_{K,\sigma}^n)^- + S_K^n.$$

Then we get $e_K^{n+1} > 0$ under the following CFL condition:

$$\delta t \le \frac{|K|}{\gamma\ \sum_{\sigma\in\mathcal{E}(K)} |\sigma|(u_{K,\sigma}^n)^+}.$$

Let us now derive a condition for the non-negativity of the source term. Applying Young's inequality to the last term of $S_{K,i}^{n+1}$, denoted by $(S_{K,i}^{n+1})_3$, we obtain

$$(S_{K,i}^{n+1})_3 \ge -\Big[\sum_{\substack{\epsilon\in\tilde{\mathcal{E}}_{\mathcal{S}}^{(i)},\ \epsilon\cap\bar{K}\neq\emptyset,\\ \epsilon=D_\sigma|D_{\sigma'},\, F_{\sigma,\epsilon}^n\le 0}} \alpha_{K,\epsilon}\,\frac{|F_{\sigma,\epsilon}^n|}{2}\Big]\,\big(u_{\sigma,i}^{n+1} - u_{\sigma,i}^n\big)^2 - \sum_{\substack{\epsilon\in\tilde{\mathcal{E}}_{\mathcal{S}}^{(i)},\ \epsilon\cap\bar{K}\neq\emptyset,\\ \epsilon=D_\sigma|D_{\sigma'},\, F_{\sigma,\epsilon}^n\le 0}} \alpha_{K,\epsilon}\,\frac{|F_{\sigma,\epsilon}^n|}{2}\,(u_{\sigma',i}^n - u_{\sigma,i}^n)^2.$$

Gathering all terms of $S_{K,i}^{n+1}$ yields:

$$S_{K,i}^{n+1} \ge \sum_{\sigma\in\mathcal{E}(K)} \frac{1}{2}\big(u_{\sigma,i}^{n+1} - u_{\sigma,i}^n\big)^2\Big[\frac{|D_{K,\sigma}|}{\delta t}\rho_K^{n+1} - \sum_{\epsilon\in\tilde{\mathcal{E}}(D_\sigma),\ \epsilon\cap\bar{K}\neq\emptyset} \alpha_{K,\epsilon}\,(F_{\sigma,\epsilon}^n)^-\Big],$$

thus $S_{K,i}^{n+1}$ is non-negative under the CFL condition:

$$\delta t \le \frac{|D_{K,\sigma}|\ \rho_K^{n+1}}{\displaystyle\sum_{\epsilon\in\tilde{\mathcal{E}}(D_\sigma),\ \epsilon\cap\bar{K}\neq\emptyset} \alpha_{K,\epsilon}\,(F_{\sigma,\epsilon}^n)^-}, \qquad \forall\sigma\in\mathcal{E}(K),$$

which concludes the proof. □

5. Passing to the limit in the scheme. The objective of this section is to show, in the one dimensional case, that if a sequence of solutions is controlled in suitable norms and converges to a limit, this latter necessarily satisfies a (part of the) weak formulation of the continuous problem.

The 1D version of the scheme which is studied in this section may be obtained from Scheme (3.1) by taking the MAC variant of the scheme, using only one horizontal stripe of grid cells, supposing that the vertical component of the velocity (the degrees of freedom of which are located on the top and bottom boundaries) vanishes, and that the measure of the vertical faces is equal to 1. For the sake of readability, however, we completely rewrite this 1D scheme, and, to this purpose, we first introduce some adaptations of the notations to the one dimensional case. For any face $\sigma \in \mathcal{E}$, let x_σ be its abscissa. For $K \in \mathcal{M}$, we denote by h_K its length (so $h_K = |K|$); when we write $K = [\sigma\sigma']$, this means that either $K = (x_\sigma, x_{\sigma'})$ or $K = (x_{\sigma'}, x_\sigma)$; if we need to specify the order, *i.e.* $K = (x_\sigma, x_{\sigma'})$ with $x_\sigma < x_{\sigma'}$, then we write $K = [\overrightarrow{\sigma\sigma'}]$. For an interface $\sigma = K|L$ between two cells K and L, we define $h_\sigma = (h_K + h_L)/2$, so, by definition of the dual mesh, $h_\sigma = |D_\sigma|$. If we need to specify the order of the cells K and L, say K is left of L, then we write $\sigma = \overrightarrow{K|L}$. With these notations, the explicit scheme (3.1) may be written as follows in the one dimensional setting:

$$\begin{aligned} &\forall K \in \mathcal{M}, \qquad \rho_K^0 = \frac{1}{|K|}\int_K \rho_0(x)\,\mathrm{d}x, \qquad e_K^0 = \frac{1}{|K|}\int_K e_0(x)\,\mathrm{d}x, \\ &\forall \sigma \in \mathcal{E}_{\rm int}, \qquad u_\sigma^0 = \frac{1}{|D_\sigma|}\int_{D_\sigma} u_0(x)\,\mathrm{d}x, \end{aligned} \tag{5.1a}$$

$$\forall K = [\overrightarrow{\sigma\sigma'}] \in \mathcal{M}, \quad \frac{|K|}{\delta t}(\rho_K^{n+1} - \rho_K^n) + F_{\sigma'}^n - F_\sigma^n = 0, \tag{5.1b}$$

$$\forall K = [\overrightarrow{\sigma\sigma'}] \in \mathcal{M}, \quad \frac{|K|}{\delta t}(\rho_K^{n+1}e_K^{n+1} - \rho_K^n e_K^n) + F_{\sigma'}^n e_{\sigma'}^n - F_\sigma^n e_\sigma^n + p_K^n(u_{\sigma'}^n - u_\sigma^n) = S_K^n, \tag{5.1c}$$

$$\forall K \in \mathcal{M}, \quad p_K^{n+1} = (\gamma - 1)\ \rho_K^{n+1}\ e_K^{n+1}, \tag{5.1d}$$

$$\forall \sigma = \overrightarrow{K|L} \in \mathcal{E}_{\rm int}, \quad \frac{|D_\sigma|}{\delta t}(\rho_{D_\sigma}^{n+1}u_\sigma^{n+1} - \rho_{D_\sigma}^n u_\sigma^n) + F_L^n u_L^n - F_K^n u_K^n + p_L^{n+1} - p_K^{n+1} = 0. \tag{5.1e}$$

The mass flux in the discrete mass balance equation is given, for $\sigma \in \mathcal{E}_{\rm int}$, by $F_\sigma^n = \rho_\sigma^n u_\sigma^n$, where the upwind approximation for the density at the face, ρ_σ^n, is defined by (3.4).

In the convection terms of the internal energy balance, the approximation for e_σ^n is upwind with respect to F_σ^n (*i.e.*, for $\sigma = \overrightarrow{K|L} \in \mathcal{E}_{\rm int}$, $e_\sigma^n = e_K^n$ if $F_\sigma^n \geq 0$ and $e_\sigma^n = e_L^n$ otherwise). The corrective term S_K^n reads, for $1 \leq n \leq N$ and $\forall K = [\sigma' \to \sigma]$:

$$\begin{aligned} S_K^n = \frac{|K|}{4\,\delta t}\,\rho_K^n\left[(u_\sigma^n - u_\sigma^{n-1})^2 + (u_{\sigma'}^n - u_{\sigma'}^{n-1})^2\right] + \frac{|F_K^{n-1}|}{2}\,(u_\sigma^{n-1} - u_{\sigma'}^{n-1})^2 \\ - |F_K^{n-1}|(u_\sigma^n - u_\sigma^{n-1})\,(u_{\sigma'}^{n-1} - u_\sigma^{n-1}), \end{aligned} \tag{5.2}$$

where the notation $K = [\sigma' \to \sigma]$ means that the flow goes from σ' to σ (*i.e.*, if $F_K^n \geq 0$, $K = [\overrightarrow{\sigma'\sigma}]$ and, if $F_K^n \leq 0$, $K = [\overrightarrow{\sigma\sigma'}]$). At the first time step, we set $S_K^0 = 0$, $\forall K \in \mathcal{M}$.

In the momentum balance equation, the application of the procedure described in Section 3 yields for the density associated to the dual cell D_σ with $\sigma = K|L$ and for the mass fluxes at the dual face located at the center of the mesh $K = [\overrightarrow{\sigma\sigma'}]$:

$$\text{for } k = n \text{ and } k = n+1,\ \rho_{D_\sigma}^k = \frac{1}{2\,|D_\sigma|}\,(|K|\,\rho_K^k + |L|\,\rho_L^k), \qquad F_K^n = \frac{1}{2}\,(F_\sigma^n + F_{\sigma'}^n), \tag{5.3}$$

and the approximation of the velocity at this face is upwind: $u_K^n = u_\sigma^n$ if $F_K^n \geq 0$ and $u_K^n = u_{\sigma'}^n$ otherwise.

Let a sequence of discretizations $(\mathcal{M}^{(m)},\ \delta t^{(m)})_{m\in\mathbb{N}}$ be given. We define the size $h^{(m)}$ of the mesh $\mathcal{M}^{(m)}$ by $h^{(m)} = \sup_{K\in\mathcal{M}^{(m)}} h_K$. Let $\rho^{(m)}$, $p^{(m)}$, $e^{(m)}$ and $u^{(m)}$ be the solution given by the scheme (5.1) with the mesh $\mathcal{M}^{(m)}$ and the time step $\delta t^{(m)}$. To the discrete unknowns, we associate piecewise constant functions on time intervals and on primal or dual meshes, so the density $\rho^{(m)}$, the pressure $p^{(m)}$, the internal energy $e^{(m)}$ and the velocity $u^{(m)}$ are defined almost everywhere on $\Omega\times(0,T)$ by:

$$\begin{aligned}
\rho^{(m)}(x,t) &= \sum_{n=0}^{N-1}\sum_{K\in\mathcal{M}} (\rho^{(m)})_K^n\, \mathcal{X}_K(x)\, \mathcal{X}_{[n,n+1)}(t),\\
p^{(m)}(x,t) &= \sum_{n=0}^{N-1}\sum_{K\in\mathcal{M}} (p^{(m)})_K^n\, \mathcal{X}_K(x)\, \mathcal{X}_{[n,n+1)}(t),\\
e^{(m)}(x,t) &= \sum_{n=0}^{N-1}\sum_{K\in\mathcal{M}} (e^{(m)})_K^n\, \mathcal{X}_K(x)\, \mathcal{X}_{[n,n+1)}(t),\\
u^{(m)}(x,t) &= \sum_{n=0}^{N-1}\sum_{\sigma\in\mathcal{E}} (u^{(m)})_\sigma^n\, \mathcal{X}_{D_\sigma}(x)\, \mathcal{X}_{[n,n+1)}(t),
\end{aligned} \tag{5.4}$$

where $\mathcal{X}_K$, $\mathcal{X}_{D_\sigma}$ and $\mathcal{X}_{[n,n+1)}$ stand for the characteristic function of the intervals K, D_σ and $[t^n, t^{n+1})$ respectively.

For discrete functions q and v defined on the primal and dual mesh respectively, we define a discrete $\mathrm{L}^1((0,T);\mathrm{BV}(\Omega))$ norm by:

$$\|q\|_{\mathcal{T},x,\mathrm{BV}} = \sum_{n=0}^{N}\delta t \sum_{\sigma=K|L\in\mathcal{E}_{\mathrm{int}}} |q_L^n - q_K^n|, \quad \|v\|_{\mathcal{T},x,\mathrm{BV}} = \sum_{n=0}^{N}\delta t \sum_{\epsilon=D_\sigma|D_{\sigma'}\in\tilde{\mathcal{E}}_{\mathrm{int}}} |v_{\sigma'}^n - v_\sigma^n|,$$

and a discrete $\mathrm{L}^1(\Omega;\mathrm{BV}((0,T)))$ norm by:

$$\|q\|_{\mathcal{T},t,\mathrm{BV}} = \sum_{K\in\mathcal{M}} |K| \sum_{n=0}^{N-1} |q_K^{n+1} - q_K^n|, \quad \|v\|_{\mathcal{T},t,\mathrm{BV}} = \sum_{\sigma\in\mathcal{E}} |D_\sigma| \sum_{n=0}^{N-1} |v_\sigma^{n+1} - v_\sigma^n|.$$

For the consistency result that we are seeking (Theorem 5.2 below), we have to assume that a sequence of discrete solutions $(\rho^{(m)}, p^{(m)}, e^{(m)}, u^{(m)})_{m\in\mathbb{N}}$ satisfies $\rho^{(m)} > 0$, $p^{(m)} > 0$ and $e^{(m)} > 0$, $\forall m\in\mathbb{N}$ (which may be a consequence of the fact that the CFL stability condition (3.12) is satisfied), and is uniformly bounded in $\mathrm{L}^\infty(\Omega\times(0,T))^4$, *i.e.*, for $m\in\mathbb{N}$ and $0\le n\le N^{(m)}$:

$$0 < (\rho^{(m)})_K^n \le C,\ 0 < (p^{(m)})_K^n \le C,\ 0 < (e^{(m)})_K^n \le C, \quad \forall K\in\mathcal{M}^{(m)}, \tag{5.5}$$

and

$$|(u^{(m)})_\sigma^n| \le C, \quad \forall \sigma\in\mathcal{E}^{(m)}, \tag{5.6}$$

where C is a positive real number. Note that, by definition of the initial conditions of the scheme, these inequalities imply that the functions ρ_0, e_0 and u_0 belong to $\mathrm{L}^\infty(\Omega)$. We also have to assume that a sequence of discrete solutions satisfies the following uniform bounds with respect to the discrete BV-norms:

$$\|\rho^{(m)}\|_{\mathcal{T},x,\mathrm{BV}} + \|p^{(m)}\|_{\mathcal{T},x,\mathrm{BV}} + \|e^{(m)}\|_{\mathcal{T},x,\mathrm{BV}} + \|u^{(m)}\|_{\mathcal{T},x,\mathrm{BV}} \le C, \quad \forall m\in\mathbb{N}. \tag{5.7}$$

and:

$$\|u^{(m)}\|_{\mathcal{T},t,\mathrm{BV}} \le C, \quad \forall m\in\mathbb{N}. \tag{5.8}$$

We are not able to prove the estimates (5.5)–(5.8) for the solutions of the scheme; however, such inequalities are satisfied by the "interpolates" (for instance, by taking

the cell average) of the solution to a Riemann problem, and are observed in computations (of course, as far as possible, *i.e.* with a limited sequence of meshes and time steps).

A weak solution to the continuous problem satisfies, for any $\varphi \in \mathrm{C}_c^\infty\big(\Omega \times [0,T)\big)$:

$$-\int_0^T \int_\Omega \Big[\rho\,\partial_t\varphi + \rho\,u\,\partial_x\varphi\Big]\,\mathrm{d}x\,\mathrm{d}t - \int_\Omega \rho_0(x)\,\varphi(x,0)\,\mathrm{d}x = 0, \tag{5.9a}$$

$$-\int_0^T \int_\Omega \Big[\rho\,u\,\partial_t\varphi + (\rho\,u^2+p)\,\partial_x\varphi\Big]\,\mathrm{d}x\,\mathrm{d}t - \int_\Omega \rho_0(x)\,u_0(x)\,\varphi(x,0)\,\mathrm{d}x = 0, \tag{5.9b}$$

$$-\int_{\Omega\times(0,T)} \Big[\rho\,E\,\partial_t\varphi + (\rho\,E+p)\,u\,\partial_x\varphi\Big]\,\mathrm{d}x\,\mathrm{d}t - \int_\Omega \rho_0(x)\,E_0(x)\,\varphi(x,0)\,\mathrm{d}x = 0, \tag{5.9c}$$

$$p = (\gamma-1)\rho\,e, \quad E = \frac{1}{2}u^2 + e, \quad E_0 = \frac{1}{2}u_0^2 + e_0. \tag{5.9d}$$

Note that these relations are not sufficient to define a weak solution to the problem, since they do not imply anything about the boundary conditions. However, they allow to derive the Rankine-Hugoniot conditions; hence if we show that they are satisfied by the limit of a sequence of solutions to the discrete problem, this implies, loosely speaking, that *the scheme computes correct shocks* (*i.e.* shocks where the jumps of the unknowns and of the fluxes are linked to the shock speed by the Rankine-Hugoniot conditions). This is the result we are seeking and which we state in Theorem 5.2. In order to prove this theorem, we need some definitions of interpolates of regular test functions on the primal and dual mesh.

Definition 5.1 (Interpolates on one-dimensional meshes). *Let Ω be an open bounded interval of $\mathbb{R}$, let $\varphi \in \mathrm{C}_c^\infty(\Omega\times[0,T))$, and let $\mathcal{M}$ be a mesh over Ω. The interpolate $\varphi_\mathcal{M}$ of φ on the primal mesh $\mathcal{M}$ is defined by:*

$$\varphi_\mathcal{M}(x,0) = \sum_{K\in\mathcal{M}} \varphi_K^0\ \mathcal{X}_K \quad \textit{and, for } t>0,\ \varphi_\mathcal{M} = \sum_{n=0}^{N-1}\sum_{K\in\mathcal{M}} \varphi_K^{n+1}\ \mathcal{X}_K\ \mathcal{X}_{(t^n,t^{n+1}]},$$

where, for $0\le n\le N$ and $K\in\mathcal{M}$, we set $\varphi_K^n = \varphi(x_K,t^n)$, with x_K the mass center of K. The time discrete derivative of the discrete function $\varphi_\mathcal{M}$ is defined by:

$$\eth_t\varphi_\mathcal{M} = \sum_{n=0}^{N-1}\sum_{K\in\mathcal{M}} \frac{\varphi_K^{n+1}-\varphi_K^n}{\delta t}\ \mathcal{X}_K\ \mathcal{X}_{(t^n,t^{n+1}]},$$

and its space discrete derivative by:

$$\eth_x\varphi_\mathcal{M} = \sum_{n=0}^{N-1} \sum_{\sigma=\overrightarrow{K|L}\in\mathcal{E}_{\rm int}} \frac{\varphi_L^{n+1}-\varphi_K^{n+1}}{h_\sigma}\ \mathcal{X}_{D_\sigma}\ \mathcal{X}_{(t^n,t^{n+1}]}.$$

Let $\varphi_\mathcal{E}$ be an interpolate of φ on the dual mesh, defined by:

$$\varphi_\mathcal{E}(x,0) = \sum_{\sigma\in\mathcal{E}} \varphi_\sigma^0\ \mathcal{X}_{D_\sigma} \quad \textit{and, for } t>0,\ \varphi_\mathcal{E} = \sum_{n=0}^{N-1}\sum_{\sigma\in\mathcal{E}} \varphi_\sigma^{n+1}\ \mathcal{X}_{D_\sigma}\ \mathcal{X}_{(t^n,t^{n+1}]},$$

where, for $0\le n\le N$ and $\sigma\in\mathcal{E}$, we set $\varphi_\sigma^n = \varphi(x_\sigma,t^n)$. We also define the time and space discrete derivatives of this function by:

$$\eth_t\varphi_\mathcal{E} = \sum_{n=0}^{N-1}\sum_{\sigma\in\mathcal{E}} \frac{\varphi_\sigma^{n+1}-\varphi_\sigma^n}{\delta t}\ \mathcal{X}_{D_\sigma}\ \mathcal{X}_{(t^n,t^{n+1}]},$$

$$\eth_x\varphi_\mathcal{E} = \sum_{n=0}^{N-1} \sum_{K=[\overrightarrow{\sigma\sigma'}]\in\mathcal{M}} \frac{\varphi_{\sigma'}^{n+1}-\varphi_\sigma^{n+1}}{h_K}\ \mathcal{X}_K\ \mathcal{X}_{(t^n,t^{n+1}]}.$$

Finally, we define $\eth_x \varphi_{\mathcal{M},\mathcal{E}}$ *by:*

$$\eth_x \varphi_{\mathcal{M},\mathcal{E}} = \sum_{n=0}^{N-1} \sum_{K=[\overrightarrow{\sigma\sigma'}]\in\mathcal{M}} \frac{\varphi_K^{n+1} - \varphi_\sigma^{n+1}}{h_K/2}\ \mathcal{X}_{D_{K,\sigma}}\ \mathcal{X}_{(t^n,t^{n+1}]} + \frac{\varphi_{\sigma'}^{n+1} - \varphi_K^{n+1}}{h_K/2}\ \mathcal{X}_{D_{K,\sigma'}}\ \mathcal{X}_{(t^n,t^{n+1}]}.$$

We are now in position to state the following result.

THEOREM 5.2 (Consistency of the one-dimensional explicit scheme).
Let Ω *be an open bounded interval of* $\mathbb{R}$*. We suppose that the initial data satisfies* $\rho_0 \in \mathrm{L}^\infty(\Omega)$*,* $p_0 \in \mathrm{BV}(\Omega)$*,* $e_0 \in \mathrm{L}^\infty(\Omega)$ *and* $u_0 \in \mathrm{L}^\infty(\Omega)$*. Let* $(\mathcal{M}^{(m)}, \delta t^{(m)})_{m\in\mathbb{N}}$ *be a sequence of discretizations such that both the time step* $\delta t^{(m)}$ *and the size* $h^{(m)}$ *of the mesh* $\mathcal{M}^{(m)}$ *tend to zero as* $m \to \infty$*, and let* $(\rho^{(m)}, p^{(m)}, e^{(m)}, u^{(m)})_{m\in\mathbb{N}}$ *be the corresponding sequence of solutions. We suppose that this sequence satisfies the estimates* (5.5)–(5.8) *and converges in* $\mathrm{L}^r(\Omega\times(0,T))^4$*, for* $1 \le r < \infty$*, to* $(\bar\rho, \bar p, \bar e, \bar u) \in \mathrm{L}^\infty(\Omega \times (0,T))^4$*.*

Then the limit $(\bar\rho, \bar p, \bar e, \bar u)$ *satisfies the system* (5.9).

Proof. It is clear that with the assumed convergence for the sequence of solutions, the limit satisfies the equation of state. The fact that the limit satisfies the weak mass balance equation (5.9a) and the weak momentum balance equation (5.9b) is proven in [15]. There only remains to prove that (5.9c) holds, by passing to the limit in the scheme, in the internal and the kinetic energy balance equations.

Let $\varphi \in \mathrm{C}_c^\infty(\Omega\times[0,T))$. Let $m \in \mathbb{N}$, $\mathcal{M}^{(m)}$ and $\delta t^{(m)}$ be given. Dropping for short the superscript $^{(m)}$, let $\varphi_\mathcal{M}$ be the interpolate of φ on the primal mesh and let $\eth_t\varphi_\mathcal{M}$ and $\eth_x\varphi_\mathcal{M}$ be its time and space discrete derivatives in the sense of Definition 5.1. Thanks to the regularity of φ, these functions respectively converge in $\mathrm{L}^r(\Omega\times(0,T))$, for $r \ge 1$ (including $r = +\infty$), to φ, $\partial_t\varphi$ and $\partial_x\varphi$ respectively. In addition, $\varphi_\mathcal{M}(\cdot,0)$ (which, for $K \in \mathcal{M}$ and $x \in K$, is equal to $\varphi_K^0 = \varphi(x_K,0)$) converges to $\varphi(\cdot,0)$ in $\mathrm{L}^r(\Omega)$ for $r \ge 1$.

We also define $\varphi_\mathcal{E}$, $\eth_t\varphi_\mathcal{E}$ and $\eth_x\varphi_\mathcal{E}$, as, respectively, the interpolate of φ on the dual mesh and its discrete time and space derivatives, still in the sense of Definition 5.1; once again thanks to the regularity of φ, these functions converge in $\mathrm{L}^r(\Omega\times(0,T))$, for $r \ge 1$, to φ, $\partial_t\varphi$ and $\partial_x\varphi$ respectively. As for the primal mesh interpolate, the dual mesh interpolate $\varphi_\mathcal{E}(\cdot,0)$ (which, for $\sigma \in \mathcal{E}$ and $x \in D_\sigma$, is equal to $\varphi_\sigma^0 = \varphi(x_\sigma,0)$) converges to $\varphi(\cdot,0)$ in $\mathrm{L}^r(\Omega)$ for $r \ge 1$.

Since the support of φ is compact in $\Omega\times[0,T)$, for m large enough, the interpolates of φ vanish on the boundary cells and at the last time step(s); hereafter, we assume that we are in this case.

On one hand, let us multiply the one dimensional discrete internal energy balance equation (5.1c) by $\delta t\,\varphi_K^{n+1}$, and sum the result for $0 \le n \le N-1$ and $K \in \mathcal{M}$. On the other hand, let us multiply the one-dimensional version of the discrete kinetic energy balance (4.1) by $\delta t\,\varphi_\sigma^{n+1}$, and sum the result for $0 \le n \le N-1$ and $\sigma \in \mathcal{E}_{\rm int}$. Finally, adding the two obtained relations, we get:

$$T_1^{(m)} + T_2^{(m)} + T_3^{(m)} + \tilde T_1^{(m)} + \tilde T_2^{(m)} + \tilde T_3^{(m)} = S^{(m)} - \tilde R^{(m)}, \tag{5.10}$$

where:

$$T_1^{(m)} = \sum_{n=0}^{N-1} \delta t \sum_{K\in\mathcal{M}} \frac{|K|}{\delta t}\ \left[\rho_K^{n+1}\, e_K^{n+1} - \rho_K^n\, e_K^n\right]\ \varphi_K^{n+1},$$

$$T_2^{(m)} = \sum_{n=0}^{N-1} \delta t \sum_{K=[\overrightarrow{\sigma\sigma'}]\in\mathcal{M}} \left[\rho_{\sigma'}^n\, e_{\sigma'}^n\, u_{\sigma'}^n - \rho_\sigma^n\, e_\sigma^n\, u_\sigma^n\right]\ \varphi_K^{n+1},$$

$$T_3^{(m)} = \sum_{n=0}^{N-1} \delta t \sum_{K=[\overrightarrow{\sigma\sigma'}]\in\mathcal{M}} p_K^n\,(u_{\sigma'}^n - u_\sigma^n)\;\varphi_K^{n+1},$$

$$\tilde{T}_1^{(m)} = \frac{1}{2}\sum_{n=0}^{N-1} \delta t \sum_{\sigma\in\mathcal{E}_{\rm int}} \frac{|D_\sigma|}{\delta t}\big[\rho_{D_\sigma}^{n+1}(u_\sigma^{n+1})^2 - \rho_\sigma^n (u_\sigma^n)^2\big]\;\varphi_\sigma^{n+1},$$

$$\tilde{T}_2^{(m)} = \frac{1}{2}\sum_{n=0}^{N-1} \delta t \sum_{\sigma=\overrightarrow{K|L}\in\mathcal{E}_{\rm int}} \big[F_L^n\,(u_L^n)^2 - F_K^n\,(u_K^n)^2\big]\;\varphi_\sigma^{n+1},$$

$$\tilde{T}_3^{(m)} = \sum_{n=0}^{N-1} \delta t \sum_{\sigma=\overrightarrow{K|L}\in\mathcal{E}_{\rm int}} (p_L^{n+1} - p_K^{n+1})\,u_\sigma^{n+1}\,\varphi_\sigma^{n+1},$$

$$S^{(m)} = \sum_{n=0}^{N-1} \delta t \sum_{K\in\mathcal{M}} S_K^n\;\varphi_K^{n+1}, \qquad \tilde{R}^{(m)} = \sum_{n=0}^{N-1} \delta t \sum_{\sigma\in\mathcal{E}_{\rm int}} R_\sigma^{n+1}\;\varphi_\sigma^{n+1},$$

and the quantities S_K^n and R_σ^{n+1} are given by Equation (5.2) and (the 1D version of) Equation (4.2) respectively.

Reordering the sums in $T_1^{(m)}$ yields:

$$T_1^{(m)} = -\sum_{n=0}^{N-1} \delta t \sum_{K\in\mathcal{M}} |K|\,\rho_K^n\,e_K^n\;\frac{\varphi_K^{n+1}-\varphi_K^n}{\delta t} - \sum_{K\in\mathcal{M}} |K|\,\rho_K^0\,e_K^0\;\varphi_K^0,$$

so that:

$$T_1^{(m)} = -\int_0^T\int_\Omega \rho^{(m)}\,e^{(m)}\eth_t\varphi_\mathcal{M}\,\mathrm{d}x\,\mathrm{d}t - \int_\Omega (\rho^{(m)})^0(x)\;(e^{(m)})^0(x)\;\varphi_\mathcal{M}(x,0)\,\mathrm{d}x.$$

The boundedness of ρ_0, e_0 and the definition (5.1a) of the initial conditions for the scheme ensures that the sequences $((\rho^{(m)})^0)_{m\in\mathbb{N}}$ and $((e^{(m)})^0)_{m\in\mathbb{N}}$ converge to ρ_0 and e_0 respectively in $\mathrm{L}^r(\Omega)$ for $r\geq 1$. Since, by assumption, the sequence of discrete solutions and of the interpolate time derivatives converge in $\mathrm{L}^r\big(\Omega\times(0,T)\big)$ for $r\geq 1$, we thus obtain:

$$\lim_{m\to+\infty} T_1^{(m)} = -\int_0^T\int_\Omega \bar\rho\,\bar e\,\partial_t\varphi\,\mathrm{d}x\,\mathrm{d}t - \int_\Omega \rho_0(x)\;e_0(x)\;\varphi(x,0)\,\mathrm{d}x.$$

Reordering the sums in $T_2^{(m)}$, we get:

$$T_2^{(m)} = -\sum_{n=0}^{N-1} \delta t \sum_{\sigma=\overrightarrow{K|L}\in\mathcal{E}} \rho_\sigma^n\,e_\sigma^n\,u_\sigma^n\;(\varphi_L^{n+1}-\varphi_K^{n+1}).$$

Using the fact that $h_\sigma = |D_\sigma|$, this relation reads:

$$T_2^{(m)} = -\sum_{n=0}^{N-1} \delta t \sum_{\sigma=\overrightarrow{K|L}\in\mathcal{E}} |D_\sigma|\;\rho_\sigma^n\,e_\sigma^n\,u_\sigma^n\;\frac{\varphi_L^{n+1}-\varphi_K^{n+1}}{h_\sigma},$$

thus $T_2^{(m)} = \mathcal{T}_2^{(m)} + \mathcal{R}_2^{(m)}$ with:

$$\mathcal{T}_2^{(m)} = -\sum_{n=0}^{N-1} \delta t \sum_{\sigma=\overrightarrow{K|L}\in\mathcal{E}} \Big[|D_{K,\sigma}|\;\rho_K^n\,e_K^n + |D_{L,\sigma}|\;\rho_L^n\,e_L^n\Big]\,u_\sigma^n\;\frac{\varphi_L^{n+1}-\varphi_K^{n+1}}{h_\sigma},$$

$$\mathcal{R}_2^{(m)} = -\sum_{n=0}^{N-1} \delta t \sum_{\sigma=\overrightarrow{K|L}\in\mathcal{E}} \Big[|D_\sigma|\;\rho_\sigma^n\,e_\sigma^n - |D_{K,\sigma}|\;\rho_K^n\,e_K^n - |D_{L,\sigma}|\;\rho_L^n\,e_L^n\Big]$$
$$u_\sigma^n\;\frac{\varphi_L^{n+1}-\varphi_K^{n+1}}{h_\sigma}.$$

The first expression reads:

$$\mathcal{T}_2^{(m)} = -\int_0^T \int_\Omega \rho^{(m)}\, e^{(m)}\, u^{(m)}\ \eth_x \varphi_{\mathcal{M}}\, \mathrm{d}x\, \mathrm{d}t,$$

and thus, thanks to the convergence assumptions:

$$\lim_{m\to+\infty} \mathcal{T}_2^{(m)} = -\int_0^T \int_\Omega \bar\rho\, \bar e\, \bar u\ \partial_x \varphi\, \mathrm{d}x\, \mathrm{d}t.$$

Let us make a change of notation for the orientation of σ in such a way that $\rho_\sigma^n = \rho_K^n$ and $e_\sigma^n = e_K^n$ (in other words, we choose to call K the upwind cell to σ instead of the left cell, which we denote by $\sigma = K \to L$). We thus get, with $C_\varphi = \|\partial_x \varphi\|_{\mathrm{L}^\infty(\Omega\times(0,T))}$:

$$|\mathcal{R}_2^{(m)}| \le C_\varphi \sum_{n=0}^{N-1} \delta t \sum_{\sigma=K\to L\in\mathcal{E}} |D_{L,\sigma}|\ \left|\rho_K^n\, e_K^n - \rho_L^n\, e_L^n\right|\ |u_\sigma^n|.$$

Applying the identity $2\,(ab-cd) = (a-c)(b+d) + (a+c)(b-d)$, which holds for any $\{a,b,c,d\} \subset \mathbb{R}$, to the quantity $\rho_K^n\, e_K^n - \rho_L^n\, e_L^n$, we obtain:

$$|\mathcal{R}_2^{(m)}| \le C_\varphi\ h^{(m)}\ \|u^{(m)}\|_{\mathrm{L}^\infty(\Omega\times(0,T))} \Big[\|\rho^{(m)}\|_{\mathrm{L}^\infty(\Omega\times(0,T))}\ \|e^{(m)}\|_{\mathcal{T},x,\mathrm{BV}} + \|e^{(m)}\|_{\mathrm{L}^\infty(\Omega\times(0,T))}\ \|\rho^{(m)}\|_{\mathcal{T},x,\mathrm{BV}}\Big],$$

and thus $|\mathcal{R}_2^{(m)}|$ tends to zero when m tends to $+\infty$.

For the term $\tilde T_1^{(m)}$, the definition (5.3) of ρ_{D_σ} and a reordering in the summation yield:

$$\tilde T_1^{(m)} = -\sum_{n=0}^{N-1} \delta t \sum_{\sigma=K|L\in\mathcal{E}} \Big[|D_{K,\sigma}|\,\rho_K^n + |D_{L,\sigma}|\,\rho_L^n\Big]\, u_\sigma^n\ \frac{\varphi_K^{n+1} - \varphi_K^n}{\delta t} - \sum_{\sigma=K|L\in\mathcal{E}} \Big[|D_{K,\sigma}|\,\rho_K^0 + |D_{L,\sigma}|\,\rho_L^0\Big]\, u_\sigma^0\ \varphi_K^0,$$

so that, by similar arguments as for the term $T_1^{(m)}$, we get:

$$\lim_{m\to+\infty} \tilde T_1^{(m)} = -\int_0^T \int_\Omega \frac{1}{2}\ \bar\rho\ \bar u^2\ \partial_t \varphi\, \mathrm{d}x\, \mathrm{d}t - \int_\Omega \frac{1}{2}\ \rho_0(x)\ u_0(x)^2\ \varphi(x,0)\, \mathrm{d}x.$$

Let us now turn to the term $\tilde T_2^{(m)}$. Reordering the sums, we get:

$$\tilde T_2^{(m)} = -\frac{1}{2}\sum_{n=0}^{N-1} \delta t \sum_{K=[\overrightarrow{\sigma\sigma'}]\in\mathcal{M}} F_K^n\,(u_K^n)^2\ (\varphi_{\sigma'}^{n+1} - \varphi_\sigma^{n+1}),$$

and, by definition of the mass flux at the dual edges:

$$\tilde T_2^{(m)} = -\frac{1}{4}\sum_{n=0}^{N-1} \delta t \sum_{K=[\overrightarrow{\sigma\sigma'}]\in\mathcal{M}} (\rho_\sigma^n u_\sigma^n + \rho_{\sigma'}^n u_{\sigma'}^n)\,(u_K^n)^2\ (\varphi_{\sigma'}^{n+1} - \varphi_\sigma^{n+1}),$$

where we recall that u_K^n is equal to either u_σ^n or $u_{\sigma'}^n$, depending on the sign of F_K^n. Let us write $\tilde T_2^{(m)} = \tilde{\mathcal{T}}_2^{(m)} + \tilde{\mathcal{R}}_2^{(m)}$, with:

$$\tilde{\mathcal{T}}_2^{(m)} = -\frac{1}{4}\sum_{n=0}^{N-1} \delta t \sum_{K=[\overrightarrow{\sigma\sigma'}]\in\mathcal{M}} \rho_K^n\,\left[(u_\sigma^n)^3 + (u_{\sigma'}^n)^3\right]\ (\varphi_{\sigma'}^{n+1} - \varphi_\sigma^{n+1}).$$

We have:

$$\tilde{\mathcal{T}}_2^{(m)} = -\int_0^T \int_\Omega \frac{1}{2}\ \rho^{(m)}\ (u^{(m)})^3\ \eth_x \varphi_{\mathcal{E}}\,\mathrm{d}x\,\mathrm{d}t,$$

and hence:

$$\lim_{m\to+\infty} \tilde{\mathcal{T}}_2^{(m)} = -\int_0^T \int_\Omega \frac{1}{2}\ \bar{\rho}\ \bar{u}^3\ \partial_x \varphi\,\mathrm{d}x\,\mathrm{d}t.$$

The remainder term reads:

$$\tilde{\mathcal{R}}_2^{(m)} = -\frac{1}{4}\sum_{n=0}^{N-1}\delta t \sum_{K=[\overrightarrow{\sigma\sigma'}]\in\mathcal{M}} \Big[(\rho_\sigma^n u_\sigma^n + \rho_{\sigma'}^n u_{\sigma'}^n)\,(u_K^n)^2 - \rho_K^n\left((u_\sigma^n)^3 + (u_{\sigma'}^n)^3\right)\Big] \\ (\varphi_{\sigma'}^{n+1} - \varphi_\sigma^{n+1}).$$

Using the notation $K = \sigma \to \sigma'$ in the above summation in order to have $u_K^n = u_\sigma^n$, we obtain:

$$\tilde{\mathcal{R}}_2^{(m)} = -\frac{\varepsilon}{4}\sum_{n=0}^{N-1}\delta t \sum_{K=\sigma\to\sigma'\in\mathcal{M}} \Big[(\rho_\sigma^n u_\sigma^n + \rho_{\sigma'}^n u_{\sigma'}^n)\,(u_\sigma^n)^2 - \rho_K^n\left((u_\sigma^n)^3 + (u_{\sigma'}^n)^3\right)\Big] \\ (\varphi_{\sigma'}^{n+1} - \varphi_\sigma^{n+1}).$$

Since, for $0 \le n \le N-1$ and $K \in \mathcal{M}$,

$$(\rho_\sigma^n u_\sigma^n + \rho_{\sigma'}^n u_{\sigma'}^n)\,(u_\sigma^n)^2 - \rho_K^n\left((u_\sigma^n)^3 + (u_{\sigma'}^n)^3\right) = \\ -(\rho_K^n - \rho_\sigma^n)\,(u_\sigma^n)^3 + \rho_K^n\,u_{\sigma'}^n\,(u_\sigma^n + u_{\sigma'}^n)\,(u_\sigma^n - u_{\sigma'}^n) - (\rho_K^n - \rho_{\sigma'}^n)\,u_{\sigma'}^n\,(u_\sigma^n)^2,$$

we have:

$$|\tilde{\mathcal{R}}_2^{(m)}| \le C_\varphi\ h^{(m)}\ \Big[\|u^{(m)}\|_{\mathrm{L}^\infty(\Omega\times(0,T))}^3\ \|\rho\|_{\mathcal{T},x,\mathrm{BV}} \\ + \|\rho^{(m)}\|_{\mathrm{L}^\infty(\Omega\times(0,T))}\ \|u^{(m)}\|_{\mathrm{L}^\infty(\Omega\times(0,T))}^2\ \|u^{(m)}\|_{\mathcal{T},x,\mathrm{BV}}\Big],$$

where the real number C_φ only depends on φ. Hence $|\tilde{\mathcal{R}}_2^{(m)}|$ tends to zero when m tends to $+\infty$.

We now turn to $T_3^{(m)}$ and $\tilde{T}_3^{(m)}$. By a change in the notation of the time exponents, using the fact that φ_σ vanishes at the last time step(s), we get:

$$\tilde{T}_3^{(m)} = \sum_{n=1}^{N-1}\delta t \sum_{\sigma=\overrightarrow{K|L}\in\mathcal{E}_{\mathrm{int}}} (p_L^n - p_K^n)\,u_\sigma^n\,\varphi_\sigma^n = \tilde{\mathcal{T}}_3^{(m)} + \tilde{\mathcal{R}}_3^{(m)},$$

with:

$$\tilde{\mathcal{T}}_3^{(m)} = \sum_{n=0}^{N-1}\delta t \sum_{\sigma=\overrightarrow{K|L}\in\mathcal{E}_{\mathrm{int}}} (p_L^n - p_K^n)\,u_\sigma^n\,\varphi_\sigma^{n+1},$$

$$\tilde{\mathcal{R}}_3^{(m)} = -\delta t \sum_{\sigma=\overrightarrow{K|L}\in\mathcal{E}_{\mathrm{int}}} (p_L^0 - p_K^0)\,u_\sigma^0\,\varphi_\sigma^0 \\ + \sum_{n=0}^{N-1}\delta t \sum_{\sigma=\overrightarrow{K|L}\in\mathcal{E}_{\mathrm{int}}} (p_L^n - p_K^n)\,u_\sigma^n\,(\varphi_\sigma^n - \varphi_\sigma^{n+1}).$$

We have, thanks to the regularity of φ:

$$|\tilde{\mathcal{R}}_3^{(m)}| \le C_\varphi\,\delta t^{(m)}\Big[\|(u^{(m)})^0\|_{\mathrm{L}^\infty(\Omega)}\,\|(p^{(m)})^0\|_{\mathrm{BV}(\Omega)} + \|u^{(m)}\|_{\mathrm{L}^\infty(\Omega\times(0,T))}\,\|p^{(m)}\|_{\mathcal{T},x,\mathrm{BV}}\Big].$$

Therefore, invoking the regularity of the initial conditions, this term tends to zero when m tends to $+\infty$. We now have for the other terms, reordering the summations:

$$T_3^{(m)} + \tilde{\mathcal{T}}_3^{(m)} = -\sum_{n=0}^{N-1} \delta t \sum_{K=[\overrightarrow{\sigma\sigma'}]\in\mathcal{M}} p_K^n\, u_\sigma^n\, (\varphi_K^{n+1} - \varphi_\sigma^{n+1}) + p_K^n\, u_{\sigma'}^n\, (\varphi_{\sigma'}^{n+1} - \varphi_K^{n+1})$$
$$= -\int_0^T \int_\Omega p^{(m)}\ u^{(m)}\ \eth_x \varphi_{\mathcal{M},\mathcal{E}}\, \mathrm{d}x\, \mathrm{d}t.$$

Since $\eth_x \varphi_{\mathcal{M},\mathcal{E}}$ converges to $\partial_x \varphi$ in $\mathrm{L}^r(\Omega \times (0,T))$ for any $r \geq 1$, we get:

$$\lim_{m\to+\infty} (T_3^{(m)} + \tilde{\mathcal{T}}_3^{(m)}) = -\int_0^T \int_\Omega \bar{p}\ \bar{u}\ \partial_x \varphi\, \mathrm{d}x\, \mathrm{d}t.$$

Finally, it now remains to check that $\lim_{m\to+\infty}(S^{(m)} - \tilde{R}^{(m)}) = 0$. Let us write this quantity as $S^{(m)} - \tilde{R}^{(m)} = \mathcal{R}_1^{(m)} + \mathcal{R}_2^{(m)}$ where, using $S_K^0 = 0, \forall K \in \mathcal{M}$:

$$\mathcal{R}_1^{(m)} = \sum_{n=0}^{N-1} \delta t \Big[\sum_{K\in\mathcal{M}} S_K^{n+1} \varphi_K^{n+1} - \sum_{\sigma\in\mathcal{E}} R_\sigma^{n+1} \varphi_\sigma^{n+1} \Big],$$
$$\mathcal{R}_2^{(m)} = \sum_{n=1}^{N-1} \delta t \sum_{K\in\mathcal{M}} S_K^n\ (\varphi_K^{n+1} - \varphi_K^n).$$

First, we prove that $\lim_{m\to+\infty} \mathcal{R}_1^{(m)} = 0$. Gathering and reordering the sums, we obtain $\mathcal{R}_1^{(m)} = \mathcal{R}_{1,1}^{(m)} + \mathcal{R}_{1,2}^{(m)} + \mathcal{R}_{1,3}^{(m)}$ with

$$\mathcal{R}_{1,1}^{(m)} = \frac{1}{2} \sum_{n=0}^{N-1} \delta t \sum_{\sigma=K|L\in\mathcal{E}} \Big[\frac{|D_{K,\sigma}|}{\delta t} \rho_K^{n+1} (u_\sigma^{n+1} - u_\sigma^n)^2 (\varphi_K^{n+1} - \varphi_\sigma^{n+1})$$
$$+ \frac{|D_{L,\sigma}|}{\delta t} \rho_L^{n+1} (u_\sigma^{n+1} - u_\sigma^n)^2 (\varphi_L^{n+1} - \varphi_\sigma^{n+1}) \Big],$$
$$\mathcal{R}_{1,2}^{(m)} = \frac{1}{2} \sum_{n=0}^{N-1} \delta t \sum_{K\in\mathcal{M}} |F_K^n|\ (u_\sigma^n - u_{\sigma'}^n)^2\ (\varphi_K^{n+1} - \varphi_\sigma^{n+1}),$$
$$\mathcal{R}_{1,3}^{(m)} = \sum_{n=0}^{N-1} \delta t \sum_{K=[\sigma'\to\sigma]\in\mathcal{M}} |F_K^n|\ (u_{\sigma'}^n - u_\sigma^n)\ (u_\sigma^{n+1} - u_\sigma^n)\ (\varphi_K^{n+1} - \varphi_\sigma^{n+1}).$$

We thus obtain:

$$|\mathcal{R}_{1,1}^{(m)}| \leq h^{(m)}\ C_\varphi\ \|\rho^{(m)}\|_{\mathrm{L}^\infty(\Omega\times(0,T))}\ \|u^{(m)}\|_{\mathrm{L}^\infty(\Omega\times(0,T))}\ \|u^{(m)}\|_{\mathcal{T},t,\mathrm{BV}},$$

and

$$|\mathcal{R}_{1,2}^{(m)}| + |\mathcal{R}_{1,3}^{(m)}| \leq h^{(m)}\ C_\varphi\ \|\rho^{(m)}\|_{\mathrm{L}^\infty(\Omega\times(0,T))}\ \|u^{(m)}\|^2_{\mathrm{L}^\infty(\Omega\times(0,T))}\ \|u^{(m)}\|_{\mathcal{T},x,\mathrm{BV}},$$

so all these terms tend to zero. The fact that $|\mathcal{R}_2^{(m)}|$ behaves as $\delta t^{(m)}$ my be proven by similar arguments.

Gathering the limits of all terms concludes the proof. □

Remark 5.1 (On BV-stability assumptions).
The proof of Theorem 5.2 in [15] and of Theorem 5.2 shows that the scheme is consistent under a BV-stability assumption much weaker than (5.7)-(5.7)*, namely:*

$$\lim_{m\to+\infty} (h^{(m)} + \delta t^{(m)}) \Big[\|\rho^{(m)}\|_{\mathcal{T},x,\mathrm{BV}} + \|p^{(m)}\|_{\mathcal{T},x,\mathrm{BV}}$$
$$+ \|e^{(m)}\|_{\mathcal{T},x,\mathrm{BV}} + \|u^{(m)}\|_{\mathcal{T},x,\mathrm{BV}} + \|u^{(m)}\|_{\mathcal{T},t,\mathrm{BV}} \Big] = 0.$$

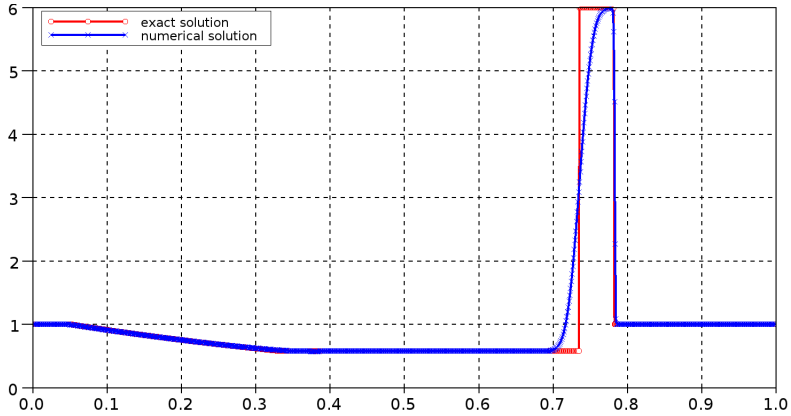

FIG. 6.1. *A Riemann problem (Test 3 of [21, Chapter 4]) – $h = 0.001$ and $\delta t = h/100$ – Density at $t = 0.012$.*

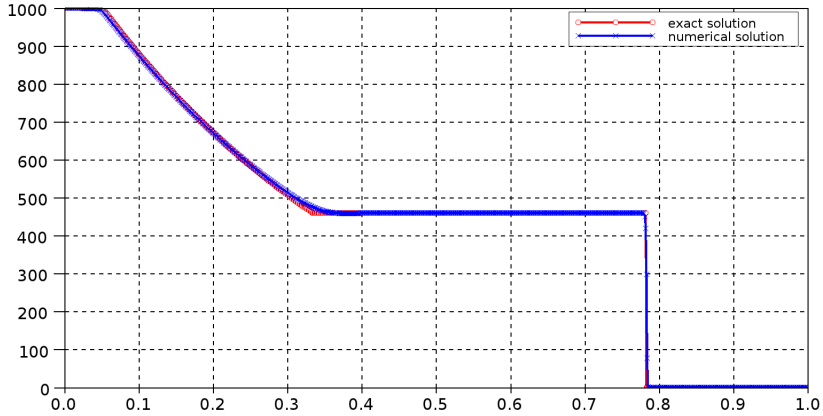

FIG. 6.2. *A Riemann problem (Test 3 of [21, Chapter 4]) – $h = 0.001$ and $\delta t = h/100$ – Pressure at $t = 0.012$.*

6. Numerical results. We assess in this section the behaviour of the scheme on a Riemann problem referred to as Test 3 in [21, Chapter 4], which is stiff enough to evidence consistency and stability properties of the scheme. The left and right states are given by:

$$\text{left state: } \begin{bmatrix} \rho_L = 1 \\ u_L = 0 \\ p_L = 1000 \end{bmatrix}; \qquad \text{right state: } \begin{bmatrix} \rho_R = 1 \\ u_R = 0 \\ p_R = 0.001 \end{bmatrix}.$$

The computational domain is $\Omega = (0, 1)$ and the final time is $T = 0.012$. The (known) analytical solution of this type of problem consists in two genuinely nonlinear waves (*i.e.* rarefaction or shock waves) separated by a contact discontinuity. For the initial data chosen in this section, the left wave is a rarefaction wave, travelling to the left, and the right wave is a shock wave, travelling to the right.

6.1. Results. The density, pressure, internal energy and velocity obtained at $t = 0.012 = T$ with $h = 0.001$ and $\delta t = h/100$ (as the maximal celerity of waves is close to 60) are shown on Figures 6.1, 6.2, 6.3 and 6.4 respectively. We observe that the scheme is rather diffusive especially for contact discontinuities for which the beneficial compressive effect of the shocks does not apply. More accurate variants may certainly be derived, using for instance MUSCL-like techniques; this work is underway.

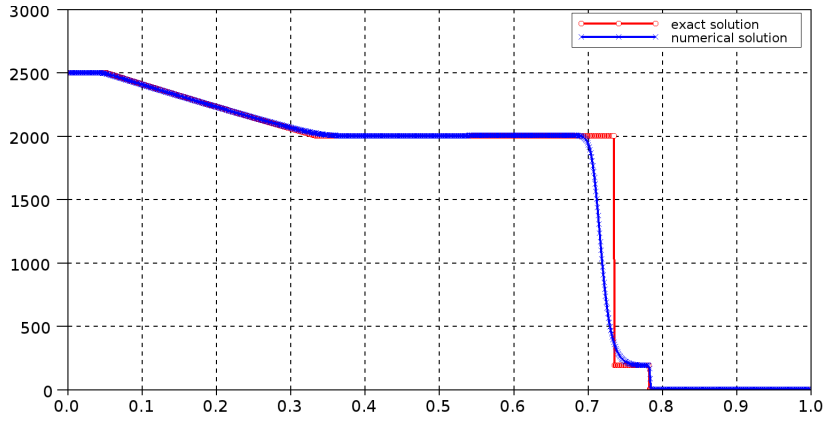

FIG. 6.3. *A Riemann problem (Test 3 of [21, Chapter 4]) – $h = 0.001$ and $\delta t = h/100$ – Internal energy at $t = 0.012$.*

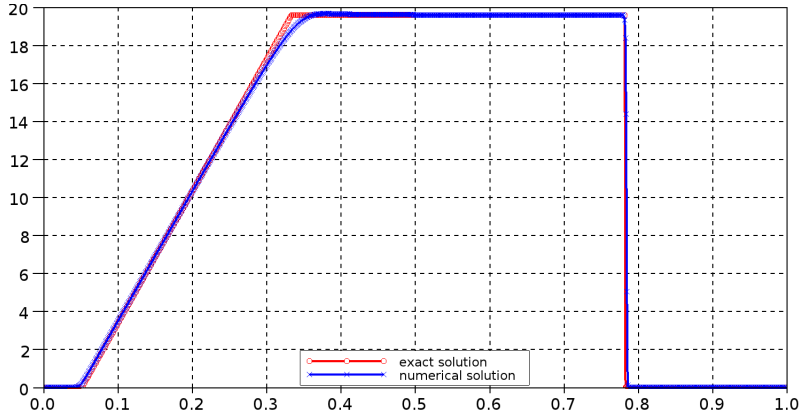

FIG. 6.4. *A Riemann problem (Test 3 of [21, Chapter 4]) – $h = 0.001$ and $\delta t = h/100$ – Velocity at $t = 0.012$.*

We also observe that the scheme keeps the velocity and pressure constant through the contact discontinuity; this may be checked directly from the expression of the discrete balance equations (precisely speaking, one may prove that, if p^n and u^n are constant, so are p^{n+1} and u^{n+1}).

In addition, we perform a convergence study, successively dividing by two the space and time steps (so keeping the CFL number constant). The differences between the computed and analytical solution at $t = 0.025$, measured in $\mathrm{L}^1(\Omega)$ norm, are reported in the following table.

space step	$h_0 = 0.001$	$h_0/2$	$h_0/4$	$h_0/8$	$h_0/16$
$\|\rho - \bar{\rho}\|_{\mathrm{L}^1(\Omega)}$	0.0651	0.0455	0.0310	0.0217	0.0153
$\|p - \bar{p}\|_{\mathrm{L}^1(\Omega)}$	1.87	1.05	0.530	0.284	0.164
$\|u - \bar{u}\|_{\mathrm{L}^1(\Omega)}$	0.0967	0.0536	0.0258	0.0134	0.00795

We measure a convergence rate which is slightly lower to 1 for the variables which are constant through the contact discontinuity (*i.e.* p and u), and equal to $1/2$ for ρ.

Finally, we test the behaviour of the scheme obtained when setting to zero the corrective terms in the internal energy balance. The density obtained with $h = 0.001$ and $\delta t = h/100$ is reported on Figure 6.5. From this result and from further numerical experiments with more and more refined meshes, it seems that the scheme converge,

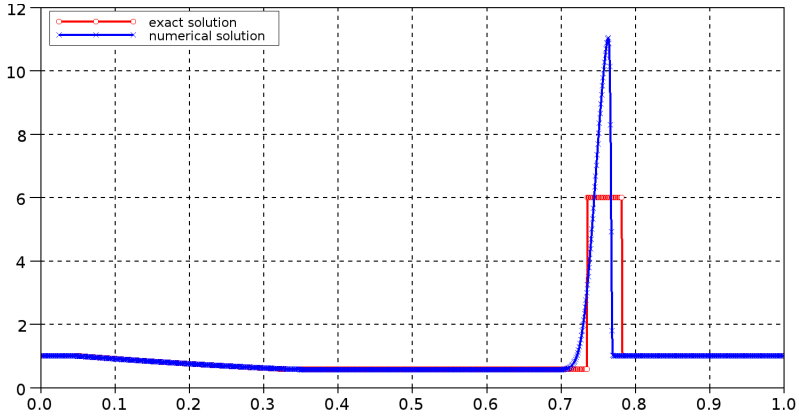

FIG. 6.5. *A Riemann problem (Test 3 of [21, Chapter 4]) – Scheme without corrective terms – $h = 0.001$ and $\delta t = h/100$ – Density at $t = 0.012$.*

but to a limit which is not a weak solution to the Euler system: indeed, the Rankine-Hugoniot condition applied to the total energy balance, with the states obtained numerically, yields a right shock velocity slightly greater than the analytical solution one, while the same shock velocity obtained numerically is clearly lower.

6.2. On a naive scheme. We also test the "naive" explicit scheme obtained by evaluating all the terms, except in time-derivative one, at time t^n. In the one dimensional setting and with the same notations as in Section 5, this scheme thus reads:

$$\forall K = [\overrightarrow{\sigma\sigma'}] \in \mathcal{M}, \quad \frac{|K|}{\delta t}(\rho_K^{n+1} - \rho_K^n) + F_{\sigma'}^n - F_\sigma^n = 0, \tag{6.1a}$$

$$\forall \sigma = \overrightarrow{K|L} \in \mathcal{E}_{\rm int},$$
$$\frac{|D_\sigma|}{\delta t}(\rho_{D_\sigma}^{n+1} u_\sigma^{n+1} - \rho_{D_\sigma}^n u_\sigma^n) + F_L^n u_L^n - F_K^n u_K^n + p_L^n - p_K^n = 0, \tag{6.1b}$$

$$\forall K = [\overrightarrow{\sigma\sigma'}] \in \mathcal{M},$$
$$\frac{|K|}{\delta t}(\rho_K^{n+1} e_K^{n+1} - \rho_K^n e_K^n) + F_{\sigma'}^n e_{\sigma'}^n - F_\sigma^n e_\sigma^n + p_K^n(u_{\sigma'}^n - u_\sigma^n) = S_K^{n+1}, \tag{6.1c}$$

$$\forall K \in \mathcal{M}, \quad p_K^{n+1} = (\gamma - 1)\ \rho_K^{n+1}\ e_K^{n+1}. \tag{6.1d}$$

Hereafter and on the figure captions, this scheme is referred to by the $\rho \rightsquigarrow u \rightsquigarrow e \rightsquigarrow p$ scheme (according to the order of update of the unknowns). Note that we are able, for this scheme also, to prove a consistency result similar to Theorem 5.2.

The computed density at time $T = 0.012$ is plotted on figures 6.9. From this result, it appears clearly that the $\rho \rightsquigarrow u \rightsquigarrow e \rightsquigarrow p$ scheme generates discontinuities in the rarefaction wave, and further experiments show that this phenomenon is not cured by a reduction of the time and space step. A similar behaviour is observed in the barotropic case (see [15] for a discussion on this issue).

7. Conclusion. We have presented in this paper an explicit scheme based on staggered meshes for Euler equations. This algorithm uses a very simple first-order upwinding strategy which consists, equation by equation, to implement an upwind discretization with respect of the material velocity of the convection term. In addition, it solves the internal energy balance instead of the total energy balance, and thus turns out to be non-conservative: indeed, the total energy conservation law is only recovered at the limit of vanishing time and space steps, thanks to the addition of corrective source terms in the discrete internal energy balance. Under CFL-like conditions based

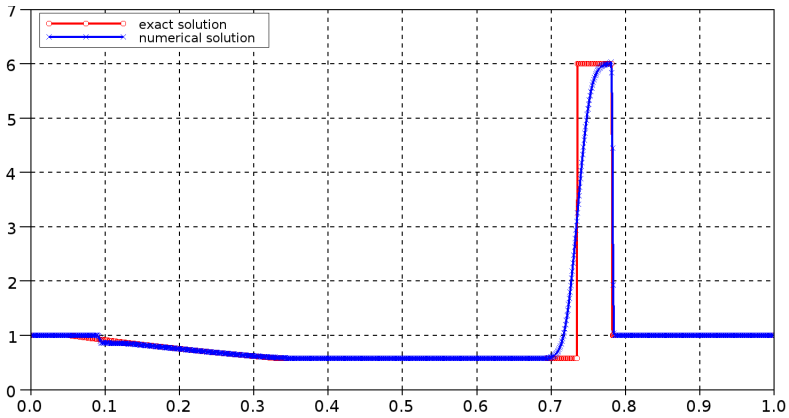

FIG. 6.6. *A Riemann problem (Test 3 of [21, Chapter 4]) – $\rho \rightsquigarrow u \rightsquigarrow e \rightsquigarrow p$ scheme – $h = 0.001$ and $\delta t = h/100$ – Density at $t = 0.012$.*

on the material velocity only (by opposition to the celerity of waves), this scheme preserves the positivity of the density, the internal energy and the pressure (in other words, the scheme preserves the convex of admissible states), and its solution satisfies a property of conservation (in fact, as often at the discrete level, non-increase) of the integral of the total energy over the computational domain. Finally, the scheme has been shown to be consistent for 1D problems, in the sense that, if a sequence of numerical solutions obtained with more and more refined meshes (and, accordingly, smaller and smaller time steps) converges, then the limit is a weak solution to the continuous problem.

This theoretical result may probably be extended in two directions: first, to check whether limits of convergent sequences are entropy solutions, and, second, to deal with the consistency issue in the multi-dimensional case. The investigation of this latter point should help to clarify the constraints on mesh generality imposed by consistency requirements, in particular with the aim to design a discretization able to cope with non-conforming locally refined meshes. This work is now being undertaken.

Numerical studies show that the proposed algorithm is stable, even if the biggest time step before blow-up is smaller than suggested by the above-mentionned CFL conditions. This behaviour had to be expected, since these CFL conditions only involve the velocity (and not the celerity of the acoutic waves): indeed, were they the only limitation, we would have obtained an explicit scheme stable up to the incompressible limit. However, the mechanisms leading to the blow-up of the scheme (or, conversely, the way to fix the time step to ensure stability) remain to be clarified, even if one may anticipate from qualitative arguments (the scheme should allow a "transport of the information" at the same speed as the continuous problem) that the time step should be small enough to avoid that the waves cross more than one mesh per time step. In addition, still as expected, the scheme is rather diffusive, especially at contact discontinuities; MUSCL-like extensions are under development to cure this problem, possibly combined with a strategy similar to the so-called entropy-viscosity technique [7, 8] to damp spurious oscillations which are sometimes observed when the velocity is small (see the companion paper [15] on the barotropic problem for a numerical study of this issue).

Since the proposed scheme uses very simple numerical fluxes, it is well suited to large multi-dimensionnal parallel computing applications, and such studies are now starting at IRSN. Still for the same reasons (and, in particular, because the construction of the discretization does not require the solution of the Riemann problem), it seems that the presented approach offers natural extensions to more complex problems, such as reacting flows; this development is foreseen at IRSN, for applications to

explosion hazards.

REFERENCES

[1] G. Ansanay-Alex, F. Babik, J.-C. Latché, and D. Vola. An L^2-stable approximation of the Navier-Stokes convection operator for low-order non-conforming finite elements. *International Journal for Numerical Methods in Fluids*, 66:555–580, 2011.

[2] F. Bouchut. *Nonlinear Stability of finite volume methods for hyperbolic conservation laws.* Birkhauser, 2004.

[3] P. G. Ciarlet. Basic error estimates for elliptic problems. In P. Ciarlet and J.L. Lions, editors, *Handbook of Numerical Analysis, Volume II*, pages 17–351. North Holland, 1991.

[4] M. Crouzeix and P.A. Raviart. Conforming and nonconforming finite element methods for solving the stationary Stokes equations. *RAIRO Série Rouge*, 7:33–75, 1973.

[5] L. Gastaldo, R. Herbin, W. Kheriji, C. Lapuerta, and J.-C. Latché. Staggered discretizations, pressure correction schemes and all speed barotropic flows. In *Finite Volumes for Complex Applications VI - Problems and Perspectives - Prague, Czech Republic*, volume 2, pages 39–56, 2011.

[6] E. Godlewski and P.-A. Raviart. *Numerical approximation of hyperbolic systems of conservation laws.* Springer, 1996.

[7] J.L Guermond and R. Pasquetti. Entropy-based nonlinear viscosity for Fourier approximations of conservation laws. *Comptes Rendus de l'Académie des Sciences de Paris – Série I – Analyse Numérique*, 346:801–806, 2008.

[8] J.L Guermond, R. Pasquetti, and B. Popov. Entropy viscosity method for nonlinear conservation laws. *Journal of Computational Physics*, 230:4248–4267, 2011.

[9] F.H. Harlow and A.A. Amsden. A numerical fluid dynamics calculation method for all flow speeds. *Journal of Computational Physics*, 8:197–213, 1971.

[10] F.H. Harlow and J.E. Welsh. Numerical calculation of time-dependent viscous incompressible flow of fluid with free surface. *Physics of Fluids*, 8:2182–2189, 1965.

[11] R. Herbin, W. Kheriji, and J.-C. Latché. Consistent semi-implicit staggered schemes for compressible flows - Part I: the barotropic Euler equations. *submitted*, 2012.

[12] R. Herbin, W. Kheriji, and J.-C. Latché. Consistent semi-implicit staggered schemes for compressible flows - Part II: the Euler equations. *submitted*, 2012.

[13] R. Herbin, W. Kheriji, and J.-C. Latché. An unconditionally stable pressure correction scheme for compressible Navier-Stokes equations. *submitted*, 2013.

[14] R. Herbin and J.-C. Latché. Kinetic energy control in the MAC discretization of the compressible Navier-Stokes equations. *International Journal of Finites Volumes*, 7, 2010.

[15] R. Herbin, J.-C. Latché, and T.T. Nguyen. Consistent explicit staggered schemes for compressible flows - Part I: the barotropic Euler equations. *submitted*, 2013.

[16] ISIS. A CFD computer code for the simulation of reactive turbulent flows. `https://gforge.irsn.fr/gf/project/isis`.

[17] B. Larrouturou. How to preserve the mass fractions positivity when computing compressible multi-component flows. *Journal of Computational Physics*, 95:59–84, 1991.

[18] M.-S. Liou. A sequel to AUSM, part II: AUSM+-up. *Journal of Computational Physics*, 214:137–170, 2006.

[19] M.-S. Liou and C.J. Steffen. A new flux splitting scheme. *Journal of Computational Physics*, 107:23–39, 1993.

[20] R. Rannacher and S. Turek. Simple nonconforming quadrilateral Stokes element. *Numerical Methods for Partial Differential Equations*, 8:97–111, 1992.

[21] E. Toro. *Riemann solvers and numerical methods for fluid dynamics – A practical introduction (third edition).* Springer, 2009.